

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2004, 15:57:50 ; Search time 36.6674 Seconds
(without alignments)
4515.528 Million cell updates/sec

Title: US-10-092-390-4
Perfect score: 3601
Sequence: 1 MVISLNSCLSFICLLCHWI.....HCDSVCAEGRWGPNC SLPCY 586

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	3601	100.0	586	5	AAE27986	Aae27986 Human EGF
2	3601	100.0	1140	5	AAE27985	Aae27985 Human EGF
3	3601	100.0	1140	7	ADD18688	Add18688 Human dis
4	3601	100.0	1192	7	ADE71305	Ade71305 Novel hum
5	1842	51.2	994	5	AAG79417	Aag79417 CADHP-6,
6	1667.5	46.3	1050	4	AAB66267	Aab66267 Human TAN
7	1466	40.7	321	4	ABG27639	Abg27639 Novel hum
8	1356.5	37.7	1350	6	ADA21141	Ada21141 Human sec
9	1313.5	36.5	1261	7	ADD78227	Add78227 Human CGD

10	1310.5	36.4	1403	6	ABJ37903	Abj37903	NOVX	prot
11	1309	36.4	878	4	ABG08033	Abg08033	Novel	hum
12	1303	36.2	1577	6	ABJ37904	Abj37904	NOVX	prot
13	1300	36.1	1398	6	ABJ37900	Abj37900	NOVX	prot
14	1300	36.1	1404	6	ABJ37899	Abj37899	NOVX	prot
15	1297.5	36.0	1450	6	ABJ37901	Abj37901	NOVX	prot
16	1272	35.3	466	4	ABG22559	Abg22559	Novel	hum
17	1252	34.8	434	4	ABB66756	Abb66756	Drosophil	
18	1247.5	34.6	1097	6	ADA21140	Ada21140	Human	sec
19	1215	33.7	636	4	AAB66269	Aab66269	Rat	TANGO
20	1203	33.4	739	6	ABU03489	Abu03489	Angiogene	
21	1169	32.5	269	4	ABG08031	Abg08031	Novel	hum
22	1169	32.5	269	6	ABO00812	Abo00812	Polypepti	
23	1169	32.5	311	6	ABO00512	Abo00512	Novel	hum
24	1167	32.4	384	4	AAG75479	Aag75479	Human	col
25	1119.5	31.1	474	4	AA72715	Aay72715	HFICU08	c
26	1002.5	27.8	587	4	ABG03826	Abg03826	Novel	hum
27	843.5	23.4	830	6	ABU79146	Abu79146	LDL	or ox
28	843.5	23.4	830	6	ABU79143	Abu79143	LDL	or ox
29	813	22.6	833	3	AA785616	Aay85616	Murine	nu
30	813	22.6	833	4	AAB60392	Aab60392	Mouse	nur
31	808	22.4	865	5	AAM47668	Aam47668	MOL8b	pro
32	808	22.4	866	4	AAB60393	Aab60393	Human	nur
33	808	22.4	866	4	AAB60394	Aab60394	Human	nur
34	808	22.4	866	5	ABG31897	Abg31897	Human	lip
35	802.5	22.3	831	5	ABB80609	Abb80609	Human	sbg
36	800	22.2	871	4	AAB60395	Aab60395	Human	nur
37	800	22.2	871	5	ABB80610	Abb80610	Human	sbg
38	785	21.8	129	4	ABG22555	Abg22555	Novel	hum
39	785	21.8	129	4	ABG27637	Abg27637	Novel	hum
40	783.5	21.8	884	5	AAM47667	Aam47667	MOL8a	pro
41	779	21.6	296	4	AAB56069	Aab56069	Skin	cell
42	779	21.6	296	5	ABB72269	Abb72269	Murine	pr
43	779	21.6	299	3	AA76014	Aay76014	Murine	EG
44	779	21.6	299	3	AA76077	Aay76077	Murine	EG
45	779	21.6	299	4	AAB55953	Aab55953	Skin	cell

ALIGNMENTS

RESULT 1

AAE27986

ID AAE27986 standard; protein; 586 AA.

XX

AC AAE27986;

XX

DT 27-JAN-2003 (first entry)

XX

DE Human EGF-family protein #2.

XX

KW Human; EGF-family protein; novel human protein; NHP; drug discovery;

KW restriction fragment length polymorphism analysis; forensic biology;

KW toxicity; infectious disease; biological disorder; medical disorder;

KW mental disorder; gene therapy.

XX

OS Homo sapiens.

XX
 PN WO200272611-A2.
 XX
 PD 19-SEP-2002.
 XX
 PF 06-MAR-2002; 2002WO-US007477.
 XX
 PR 12-MAR-2001; 2001US-0275013P.
 XX
 PA (LEXI-) LEXICON GENETICS INC.
 XX
 PI Yu X, Miranda M;
 XX
 DR WPI; 2002-723315/78.
 DR N-PSDB; AAD46319.
 XX
 PT New novel human nucleic acids useful for e.g. identifying protein coding
 PT sequences and mapping unique genes to a particular chromosome, as DNA
 PT markers for restriction fragment length polymorphism analysis, or in
 PT forensic biology.
 XX
 PS Claim 2; Page 40-42; 42pp; English.
 XX
 CC The present sequence is EGF-family protein, a novel human protein (NHP).
 CC The NHP sequences are useful for mapping unique genes to a particular
 CC chromosome; as DNA markers for restriction fragment length polymorphism
 CC analysis; in forensic biology; in defining and monitoring both drug
 CC action and toxicity; in identifying, selecting and validating novel
 CC molecular targets for drug discovery; in microarrays or other assay
 CC formats to screen collections of genetic material from patients who have
 CC a particular medical condition. The NHP peptides, fusion proteins,
 CC antibodies, antagonists and agonists can be used for detecting mutant
 CC NHPs or inappropriately expressed NHPs for the diagnosis of disease; for
 CC screening drugs for treatment of symptomatic or phenotypic manifestations
 CC of perturbing the normal function of NHP in the body and to treat
 CC diseases including infectious, mental, biological, or medical diseases or
 CC disorders. They are also used in gene therapy
 XX
 SQ Sequence 586 AA;

Query Match 100.0%; Score 3601; DB 5; Length 586;
 Best Local Similarity 100.0%; Pred. No. 2.2e-158;
 Matches 586; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVISLNSCLSFCICLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 MVISLNSCLSFCICLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC 60

 Qy 61 TDILNWFKCTRHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIA 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 61 TDILNWFKCTRHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIA 120

 Qy 121 PNTCQCEPGWGGTNCSSACGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCE 180
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 121 PNTCQCEPGWGGTNCSSACGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCE 180

 Qy 181 RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPQCEQRCPC 240

Db	181	RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPQCEQRCPC	240
Qy	241	QNGGVCHHVTGECSCPSGWMGTVCGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP	300
Db	241	QNGGVCHHVTGECSCPSGWMGTVCGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP	300
Qy	301	GYTGERCQDECPVGTYGVLCAETCQCVNNGGKCYHVSAGLCEAGFAGERCEARLCPEGLY	360
Db	301	GYTGERCQDECPVGTYGVLCAETCQCVNNGGKCYHVSAGLCEAGFAGERCEARLCPEGLY	360
Qy	361	GIKCDKRCPCHELENTSHCHPMSEGECKPGWWSGLYCNCTSPGFYGEACQICSCQNGAD	420
Db	361	GIKCDKRCPCHELENTSHCHPMSEGECKPGWWSGLYCNCTSPGFYGEACQICSCQNGAD	420
Qy	421	CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSVPDGSCTCKAGWHGV	480
Db	421	CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSVPDGSCTCKAGWHGV	480
Qy	481	DCSIRCPSTWGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAE	540
Db	481	DCSIRCPSTWGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAE	540
Qy	541	RCDCSHADGCHPTTGHCRCLPGWWSGVHCDVCAEGRWGPNCSLPCY	586
Db	541	RCDCSHADGCHPTTGHCRCLPGWWSGVHCDVCAEGRWGPNCSLPCY	586

RESULT 2

AAE27985

ID AAE27985 standard; protein; 1140 AA.

XX

AC AAE27985;

XX

DT 27-JAN-2003 (first entry)

XX

DE Human EGF-family protein #1.

XX

KW Human; EGF-family protein; novel human protein; NHP; drug discovery;
 KW restriction fragment length polymorphism analysis; forensic biology;
 KW toxicity; infectious disease; biological disorder; medical disorder;
 KW mental disorder; gene therapy.

XX

OS Homo sapiens.

XX

PN WO200272611-A2.

XX

PD 19-SEP-2002.

XX

PF 06-MAR-2002; 2002WO-US007477.

XX

PR 12-MAR-2001; 2001US-0275013P.

XX

PA (LEXI-) LEXICON GENETICS INC.

XX

PI Yu X, Miranda M;

XX

DR WPI; 2002-723315/78.
DR N-PSDB; AAD46318.

XX

PT New novel human nucleic acids useful for e.g. identifying protein coding
PT sequences and mapping unique genes to a particular chromosome, as DNA
PT markers for restriction fragment length polymorphism analysis, or in
PT forensic biology.

XX

PS Claim 2; Page 37-40; 42pp; English.

XX

CC The present sequence is EGF-family protein, a novel human protein (NHP).
CC The NHP sequences are useful for mapping unique genes to a particular
CC chromosome; as DNA markers for restriction fragment length polymorphism
CC analysis; in forensic biology; in defining and monitoring both drug
CC action and toxicity; in identifying, selecting and validating novel
CC molecular targets for drug discovery; in microarrays or other assay
CC formats to screen collections of genetic material from patients who have
CC a particular medical condition. The NHP peptides, fusion proteins,
CC antibodies, antagonists and agonists can be used for detecting mutant
CC NHPs or inappropriately expressed NHPs for the diagnosis of disease; for
CC screening drugs for treatment of symptomatic or phenotypic manifestations
CC of perturbing the normal function of NHP in the body and to treat
CC diseases including infectious, mental, biological, or medical diseases or
CC disorders. They are also used in gene therapy

XX

SQ Sequence 1140 AA;

Query Match 100.0%; Score 3601; DB 5; Length 1140;
Best Local Similarity 100.0%; Pred. No. 3.6e-158;
Matches 586; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVISLNSCLSFICLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC 60
|
Db 1 MVISLNSCLSFICLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC 60

Qy 61 TDILNWFKCTRHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRCIA 120
|
Db 61 TDILNWFKCTRHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRCIA 120

Qy 121 PNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCED 180
|
Db 121 PNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCED 180

Qy 181 RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGFPQCEQRCPC 240
|
Db 181 RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGFPQCEQRCPC 240

Qy 241 QNGGVCHHVTGECSCPSGWMGTVCQGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP 300
|
Db 241 QNGGVCHHVTGECSCPSGWMGTVCQGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP 300

Qy 301 GYTGERCQDECPVGTYGVLCAETCQCVNCGKCYHVSAGACLCEAGFAGERCEARLCPEGLY 360
|
Db 301 GYTGERCQDECPVGTYGVLCAETCQCVNCGKCYHVSAGACLCEAGFAGERCEARLCPEGLY 360

Qy 361 GIKCDKRCPCHELENTHSCHPMSEGCACKPGWSGLYCNETCSPGFYGEACQQICSCQNGAD 420
|

Db 361 GIKCDKRCPCHLENTHSCHPMSEGCACKPGWSGLYCNETCSPGFYGEACQQICSCQNGAD 420

Qy 421 CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDVCSFVDGSCCTCKAGWHGV 480
 |||

Db 421 CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDVCSFVDGSCCTCKAGWHGV 480

Qy 481 DCSIRCPSTWGFNLTCLGACNTLDGTCTCAPGWRGEKCELPQDGTGYNCAE 540
 |||

Db 481 DCSIRCPSTWGFNLTCLGACNTLDGTCTCAPGWRGEKCELPQDGTGYNCAE 540

Qy 541 RCDCHADGCHPTTGHCRLPGWSGVHCDVCAEGRWGPNCPLPCY 586
 |||

Db 541 RCDCHADGCHPTTGHCRLPGWSGVHCDVCAEGRWGPNCPLPCY 586

RESULT 3

ADD18688

ID ADD18688 standard; protein; 1140 AA.

XX

AC ADD18688;

XX

DT 15-JAN-2004 (first entry)

XX

DE Human disease related protein SeqID119.

XX

KW human; disease state; cytostatic; antiinflammatory; ophthalmological;

KW antiarteriosclerotic; vulnerary; gene therapy;

KW hypoxia-regulated condition; tumorigenesis; angiogenesis; apoptosis;

KW inflammation; erythropoiesis; glycolysis; gluconeogenesis;

KW glucose transportation; catecholamine synthesis; iron transport;

KW nitric oxide synthesis; cancer; ischaemic condition; reperfusion injury;

KW retinopathy; neonatal stress; pre-eclampsia; atherosclerosis;

KW inflammatory condition; wound healing.

XX

OS Homo sapiens.

XX

PN WO2003018621-A2.

XX

PD 06-MAR-2003.

XX

PF 23-AUG-2002; 2002WO-GB003892.

XX

PR 23-AUG-2001; 2001GB-00020558.

PR 05-OCT-2001; 2001GB-00024037.

XX

PA (OXFO-) OXFORD BIOMEDICA UK LTD.

XX

PI Kingsman SM, White J, Ward NR, Harris RA, Naylor S, Mundy CR;

XX

DR WPI; 2003-290046/28.

DR N-PSDB; ADD18689.

XX

PT New substantially purified polypeptide, useful for diagnosing or treating

PT a hypoxia-regulated condition, such as cancer, ischemia, reperfusion

PT injury, retinopathy, pre-eclampsia, atherosclerosis, inflammation, or

PT wound healing.

XX

PS Claim 25; SEQ ID NO 119; 424pp; English.

XX

CC This invention relates to novel human genes and gene product which are
CC implicated in certain disease states. Compounds which modulate the
CC proteins of the invention may have cytostatic, antiinflammatory,
CC ophthalmological, antiarteriosclerotic or vulnerary activities. The
CC sequences of the invention may be useful for gene therapy. The invention
CC may be useful for diagnosing or treating a hypoxia-regulated condition,
CC such as tumourigenesis, angiogenesis, apoptosis, inflammation,
CC erythropoiesis, or the biological response to hypoxia conditions
CC including processes such as glycolysis, gluconeogenesis, glucose
CC transportation, catecholamine synthesis, iron transport or nitric oxide
CC synthesis. The disease includes cancer, ischaemic conditions, reperfusion
CC injury, retinopathy, neonatal stress, pre-eclampsia, atherosclerosis,
CC inflammatory conditions or wound healing. The present sequence is that of
CC a disease related protein of the invention.

XX

SQ Sequence 1140 AA;

Query Match 100.0%; Score 3601; DB 7; Length 1140;
Best Local Similarity 100.0%; Pred. No. 3.6e-158;
Matches 586; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVISLNSCLSFICLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC 60
|
Db 1 MVISLNSCLSFICLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC 60

QY 61 TDILNWFKCTRHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIA 120
|
Db 61 TDILNWFKCTRHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIA 120

QY 121 PNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCED 180
|
Db 121 PNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCED 180

QY 181 RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKGHPQCEQRCPC 240
|
Db 181 RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKGHPQCEQRCPC 240

QY 241 QNGGVCHHVTGECSCPSGWMGTVCQGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP 300
|
Db 241 QNGGVCHHVTGECSCPSGWMGTVCQGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP 300

QY 301 GYTGERCQDECPVGTYGVLCAETCQCVNNGKCYHVS GACLCEAGFAGERCEARLCPEGLY 360
|
Db 301 GYTGERCQDECPVGTYGVLCAETCQCVNNGKCYHVS GACLCEAGFAGERCEARLCPEGLY 360

QY 361 GIKCDKRCPCCHLENTSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGAD 420
|
Db 361 GIKCDKRCPCCHLENTSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGAD 420

QY 421 CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSPVDGSCTCKAGWHGV 480
|
Db 421 CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSPVDGSCTCKAGWHGV 480

QY 481 DCSIRCPSGTWGFGCNLTQCCLNGGACNTLDGTCTCAPGWRGEKCELPCQDGTYGLNCAE 540
|

Db 481 DCSIRCPSGTWGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAE 540

Qy 541 RCDCSHADGCHPTTGHCRCLPGWSGVHCDSVCAEGRWGPNCSLPCY 586
 |||

Db 541 RCDCSHADGCHPTTGHCRCLPGWSGVHCDSVCAEGRWGPNCSLPCY 586

RESULT 4

ADE71305

ID ADE71305 standard; protein; 1192 AA.

XX

AC ADE71305;

XX

DT 29-JAN-2004 (first entry)

XX

DE Novel human protein #59.

XX

KW human; novel protein; drug.

XX

OS Homo sapiens.

XX

PN JP2002345493-A.

XX

PD 03-DEC-2002.

XX

PF 29-MAR-2001; 2002JP-00049046.

XX

PR 29-MAR-2001; 2001JP-00095524.

XX

PA (KAZU-) ZH KAZUSA DNA KENKYUSHO.

XX

DR WPI; 2003-460885/44.

DR N-PSDB; ADE71243.

XX

PT A gene and a protein encoded by it, used in drugs.

XX

PS Disclosure; Page 242-247; 257pp; Japanese.

XX

CC The invention comprises the amino acid and coding sequences of novel
 CC human proteins. The DNA and protein sequences of the invention are used
 CC in drugs. The present amino acid sequence represents a novel human
 CC protein of the invention.

XX

SQ Sequence 1192 AA;

Query Match 100.0%; Score 3601; DB 7; Length 1192;

Best Local Similarity 100.0%; Pred. No. 3.7e-158;

Matches 586; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVISLNSCLSFICLLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC 60
 |||

Db 53 MVISLNSCLSFICLLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC 112

Qy 61 TDILNWFKCTRHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRCIA 120
 |||

Db 113 TDILNWFKCTRHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRCIA 172

Qy	121	PNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCED	180
Db	173	PNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCED	232
Qy	181	RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKGHPQCEQRCPC	240
Db	233	RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKGHPQCEQRCPC	292
Qy	241	QNGGVCHHVTGECSCPSGWMGTVCGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP	300
Db	293	QNGGVCHHVTGECSCPSGWMGTVCGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP	352
Qy	301	GYTGERCQDECPVGTGYVLCAETCQCVNNGGKCYHVS GACLCEAGFAGERCEARLCPEGLY	360
Db	353	GYTGERCQDECPVGTGYVLCAETCQCVNNGGKCYHVS GACLCEAGFAGERCEARLCPEGLY	412
Qy	361	GIKCDKRCPCCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGAD	420
Db	413	GIKCDKRCPCCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGAD	472
Qy	421	CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCS PVDGSCTCKAGWHGV	480
Db	473	CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCS PVDGSCTCKAGWHGV	532
Qy	481	DCSIRCPSGTWGFGCNLTQCCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTGYGLNCAE	540
Db	533	DCSIRCPSGTWGFGCNLTQCCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTGYGLNCAE	592
Qy	541	RCDCSHADGCHPTTGHCRCCLPGWSGVHCDSVCAEGRWGPNC SLPCY	586
Db	593	RCDCSHADGCHPTTGHCRCCLPGWSGVHCDSVCAEGRWGPNC SLPCY	638

RESULT 5

AAG79417

ID AAG79417 standard; protein; 994 AA.

XX

AC AAG79417;

XX

DT 25-OCT-2002 (first entry)

XX

DE CADHP-6, Incyte ID No: 4097936CD1.

XX

KW Human; cell adhesion protein; CADHP; AIDS; Alzheimer's disease;

KW acquired immunodeficiency syndrome; thymic dysplasia; epilepsy;

KW renal tubular acidosis; congenital glaucoma; cancer; atherosclerosis;

KW Parkinson's disease.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Domain 1. .609

FT /label= Sushi_repeat

FT /note= "Identified by BLAST-DOMO"

FT Peptide 1. .29

FT /label= Signal_peptide

FT /note= "Identified by HMMER"

FT	Peptide	1. .28
FT		/label= Signal_peptide
FT		/note= "Identified by HMMER"
FT	Peptide	1. .25
FT		/label= Signal_peptide
FT		/note= "Identified by HMMER"
FT	Peptide	1. .24
FT		/label= Signal_peptide
FT		/note= "Identified by HMMER"
FT	Peptide	1. .22
FT		/label= Signal_peptide
FT		/note= "Identified by HMMER"
FT	Peptide	1. .20
FT		/label= Signal_peptide
FT		/note= "Identified by HMMER"
FT	Peptide	1. .20
FT		/label= Signal_cleavage
FT		/note= "Identified by SPSCAN"
FT	Peptide	1. .19
FT		/label= Signal_peptide
FT		/note= "Identified by HMMER"
FT	Peptide	1. .18
FT		/label= Signal_peptide
FT		/note= "Identified by HMMER"
FT	Peptide	1. .16
FT		/label= Signal_peptide
FT		/note= "Identified by HMMER"
FT	Modified-site	30
FT		/note= "Potentially phosphorylated"
FT	Modified-site	38
FT		/note= "Potentially phosphorylated"
FT	Domain	101. .131
FT		/label= EGF-like_domain
FT		/note= "Identified by HMMER-PFAM"
FT	Domain	120. .131
FT		/label= EGF-like_domain_signature_2
FT		/note= "Identified by MOTIFS"
FT	Domain	120. .131
FT		/label= EGF-like_domain_signature_1
FT		/note= "Identified by MOTIFS"
FT	Binding-site	127. .129
FT		/label= Cell_attachmnt_sequence
FT		/note= "Identified by MOTIFS"
FT	Peptide	133. .161
FT		/label= Type_III_EGF-like_signature
FT		/note= "Identified by BLIMPS-PRINTS"
FT	Domain	138. .576
FT		/label= Sushi_repeat
FT		/note= "Identified by BLAST-DOMO"
FT	Domain	144. .174
FT		/label= EGF-like_domain
FT		/note= "Identified by HMMER-PFAM"
FT	Modified-site	152
FT		/note= "Potentially glycosylated"
FT	Modified-site	153
FT		/note= "Potentially glycosylated"
FT	Modified-site	154

FT		/note= "Potentially phosphorylated"
FT	Domain	187. .216
FT		/label= EGF-like_domain
FT		/note= "Identified by HMMER-PFAM"
FT	Domain	205. .216
FT		/label= EGF-like_domain_signature_1
FT		/note= "Identified by MOTIFS"
FT	Domain	229. .259
FT		/label= EGF-like_domain
FT		/note= "Identified by HMMER-PFAM"
FT	Domain	248. .259
FT		/label= EGF-like_domain_signature_1
FT		/note= "Identified by MOTIFS"
FT	Domain	248. .259
FT		/label= EGF-like_domain_signature_2
FT		/note= "Identified by MOTIFS"
FT	Domain	248. .252
FT		/label= Sushi_domain_protein
FT		/note= "Identified by BLIMPS-PFAM"
FT	Modified-site	271
FT		/note= "Potentially glycosylated"
FT	Domain	272. .302
FT		/label= EGF-like_domain
FT		/note= "Identified by HMMER-PFAM"
FT	Peptide	284. .302
FT		/label= Type_III_EGF-like_signature
FT		/note= "Identified by BLIMPS-PRINTS"
FT	Domain	291. .302
FT		/label= EGF-like_domain_signature_1
FT		/note= "Identified by MOTIFS"
FT	Domain	291. .302
FT		/label= EGF-like_domain_signature_2
FT		/note= "Identified by MOTIFS"
FT	Domain	315. .345
FT		/label= EGF-like_domain
FT		/note= "Identified by HMMER-PFAM"
FT	Domain	334. .345
FT		/label= EGF-like_domain_signature_1
FT		/note= "Identified by MOTIFS"
FT	Domain	334. .345
FT		/label= EGF-like_domain_signature_2
FT		/note= "Identified by MOTIFS"
FT	Modified-site	346
FT		/note= "Potentially phosphorylated"
FT	Modified-site	355
FT		/note= "Potentially phosphorylated"
FT	Domain	365. .391
FT		/label= EGF-like_domain
FT		/note= "Identified by HMMER-PFAM"
FT	Domain	380. .391
FT		/label= EGF-like_domain_signature_1
FT		/note= "Identified by MOTIFS"
FT	Domain	380. .391
FT		/label= EGF-like_domain_signature_2
FT		/note= "Identified by MOTIFS"
FT	Modified-site	392
FT		/note= "Potentially glycosylated"

FT	Domain	404. .434
FT		/label= EGF-like_domain
FT		/note= "Identified by HMMER-PFAM"
FT	Domain	423. .434
FT		/label= EGF-like_domain_signature_1
FT		/note= "Identified by MOTIFS"
FT	Domain	423. .434
FT		/label= EGF-like_domain_signature_2
FT		/note= "Identified by MOTIFS"
FT	Modified-site	446
FT		/note= "Potentially glycosylated"
FT	Domain	447. .477
FT		/label= EGF-like_domain
FT		/note= "Identified by HMMER-PFAM"
FT	Modified-site	448
FT		/note= "Potentially phosphorylated"
FT	Modified-site	460
FT		/note= "Potentially phosphorylated"
FT	Domain	466. .477
FT		/label= EGF-like_domain_signature_2
FT		/note= "Identified by MOTIFS"
FT	Modified-site	476
FT		/note= "Potentially glycosylated"
FT	Domain	490. .520
FT		/label= EGF-like_domain
FT		/note= "Identified by HMMER-PFAM"
FT	Modified-site	491
FT		/note= "Potentially glycosylated"
FT	Domain	509. .520
FT		/label= EGF-like_domain_signature_1
FT		/note= "Identified by MOTIFS"
FT	Domain	509. .520
FT		/label= EGF-like_domain_signature_2
FT		/note= "Identified by MOTIFS"
FT	Domain	533. .563
FT		/label= EGF-like_domain
FT		/note= "Identified by HMMER-PFAM"
FT	Modified-site	535
FT		/note= "Potentially phosphorylated"
FT	Domain	552. .563
FT		/label= EGF-like_domain_signature_1
FT		/note= "Identified by MOTIFS"
FT	Domain	552. .563
FT		/label= EGF-like_domain_signature_2
FT		/note= "Identified by MOTIFS"
FT	Modified-site	566
FT		/note= "Potentially phosphorylated"
FT	Modified-site	575
FT		/note= "Potentially glycosylated"
FT	Domain	576. .606
FT		/label= EGF-like_domain
FT		/note= "Identified by HMMER-PFAM"
FT	Modified-site	581
FT		/note= "Potentially phosphorylated"
FT	Domain	595. .606
FT		/label= EGF-like_domain_signature_1
FT		/note= "Identified by MOTIFS"

AC AAB66267;
 XX
 DT 05-APR-2001 (first entry)
 XX
 DE Human TANGO 272 SEQ ID NO: 14.
 XX
 KW Membrane associated protein; secreted protein; human; mouse; rat;
 KW INTERCEPT 340; MANGO 003; MANGO 347; TANGO 272; TANGO 295; TANGO 354;
 KW TANGO 378; skeletal disorder; cardiovascular disorder; renal disorder;
 KW haematopoietic disorder; neural disorder; hepatic disorder;
 KW neoplastic disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200100673-A1.
 XX
 PD 04-JAN-2001.
 XX
 PF 29-JUN-2000; 2000WO-US018198.
 XX
 PR 30-JUN-1999; 99US-00345464.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Barnes TM, Fraser CC, Wrighton N, Myers P, Busfield SJ, Sharp JD;
 XX
 DR WPI; 2001-050128/06.
 DR N-PSDB; AAF27787.
 XX
 PT Isolated secreted or transmembrane proteins are used for diagnosis and
 PT treatment of neoplastic and hematopoietic disorders e.g. T cell
 PT disorders, cancer and tumors.
 XX
 PS Claim 9; Page 227-229; 294pp; English.
 XX
 CC The present invention provides the protein and coding sequences for a
 CC number of membrane associated and secreted proteins from human, mouse and
 CC rat. The proteins are designated INTERCEPT 340, MANGO 003, MANGO 347,
 CC TANGO 272, TANGO 295, TANGO 254 and TANGO 378. The proteins are all
 CC involved in signal transduction and the sequences can be used in the
 CC treatment of cardiovascular, renal, hepatic, neural, neoplastic, skeletal
 CC and haematopoietic disorders
 XX
 SQ Sequence 1050 AA;

Query Match 46.3%; Score 1667.5; DB 4; Length 1050;
 Best Local Similarity 45.0%; Pred. No. 2.6e-69;
 Matches 284; Conservative 61; Mismatches 181; Indels 105; Gaps 8;

Qy 14 LLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSCTDILNW----- 66
 ||| : | || || |||::| :||: || : |
 Db 9 LLLAVGLRLAGTLNPSDPNTCSFEWSEFTTTTKESHSRPFSLLPSEPCE--RPWEGPHTCP 66
 Qy 67 -----FKCTRHRVSYR-----TAY 80
 | | :| :
 Db 67 SPQTQRKLLASRDSFCMVCVGAGVQWRDRSALQPQTGNALSMRPQPRVLSGAPSLASPGH 126

QY 81 RHGEKTM YRRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGWGGTNCSSA-- 138
 Db 127 TVVVKTDHRQLQCCHGFYESRGFCVPLCAQECVHGRCVAPNQCCVPGWRGDDCSSAPN 186

QY 139 ----CDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYNDC HQR 194
 Db 187 CLQPCTPGYYGPACQFRCQC-HGAPCDPQTGACFCPAERTGPSCDVSCSQGT----- 237

QY 195 CQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGFPQCEQRCPCQNGGVCHHVTGEC S 254
 Db 238 -----SGFFC-----PSTH-----PCQNGGVFQTPQGSCS 262

QY 255 CP SGWMTVCGQPCPEGRFGKNCSQECQCHNGGTCD AATGQCHCSPGYTGERCQDECPVG 314
 Db 263 CPPGWMGTICSLPCPEGFHGPNC SQECRCHNGGLCDRFTGQCRCAPGYTGDRCREEC PVG 322

QY 315 TYGVLCAETCQCVNNGKCYHVSGACLCEAGFAGERCEARLCPEGLYGIKCDKRCPC HLEN 374
 Db 323 RFGQDCAETCDCAPDARCFPANGACLCEHGFTGDRCTDRLCPDGFYGLSCQAPCTCDREH 382

QY 375 THSCHPM SGECACKPGW SGLYCNETCSPGFYGEACQ QICSCQNGADCDSVTGKCTCAPGF 434
 Db 383 SL SCHPMNGEC SCLPGWAGLHCNESCPQDTHGPGCQEHCLCLHGGVCQAT SGLCQCAPGY 442

QY 435 KGIDCSTPCPLGTYGINCSSRCGCKNDAVCS PVDGSCTCKAGWHGVDCSIRCPSGTWGFG 494
 Db 443 TGPHCASLCPDPTYGVNCSARCS CENAIACSPIDGECVCKEGWQRGNC SVPCPPGTWGFS 502

QY 495 CNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPQDGT YGLNCAERCDCSHADGCHPTT 554
 Db 503 CNASCQCAHEAVCS PQTGACTCTPGWHGAHCQLPCPKGQFGEGCASRCD CDHSDGCDPVH 562

QY 555 GHCRC L PGWSGVHCD SVCAEGRWGPNCSLPC 585
 Db 563 GRCQCQAGWMGARCHLSCPEGLWGVNCSNTC 593

RESULT 7

ABG27639

ID ABG27639 standard; protein; 321 AA.

XX

AC ABG27639;

XX

DT 18-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #27630.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US008631.

XX
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS91826.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20; SEQ ID NO 57998; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 321 AA;

Query Match 40.7%; Score 1466; DB 4; Length 321;
 Best Local Similarity 90.3%; Pred. No. 2.1e-60;
 Matches 241; Conservative 0; Mismatches 0; Indels 26; Gaps 2;

Qy 137 SACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQ 196
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 9 SACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQ 68
 Qy 197 CQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPQCEQRCPCQNGGVCHHVTGECSCP 256
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 69 CQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPQCEQRCPCQNGGVCHHVTGECSCP 128
 Qy 257 SGWM-----GTVCGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHC 298
 |||| |
 Db 129 SGWMLSFPGWRPIXFSKSLXMQGTVCGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHC 188

Qy 299 SPGYTGER-----CQDECPVGTYGVLCAETCQCVNGGKCYHVSAGACLCEAGFAGERC 350
 ||||| |||||
 Db 189 SPGYTGERAAVPDVRKCQDECPVGTYGVLCAETCQCVNGGKCYHVSAGACLCEAGFAGERC 248
 Qy 351 EARLCPEGLYGIKCDKRCPCCHLENTHS 377
 |||||
 Db 249 EARLCPEGLYGIKCDKRCPCCHLENTHS 275

RESULT 8

ADA21141

ID ADA21141 standard; protein; 1350 AA.

XX

AC ADA21141;

XX

DT 20-NOV-2003 (first entry)

XX

DE Human secreted protein SECP-46 SEQ ID NO:46.

XX

KW human; secreted protein; SECP; anti-HIV; antiallergic; antiinflammatory;
 KW antianaemic; antiparkinsonian; nootropic; anticonvulsant;
 KW antiarteriosclerotic; antiasthmatic; immunosuppressive; antithyroid;
 KW cytostatic; hepatotropic; dermatological; antidiabetic; nephrotropic;
 KW antigout; thyromimetic; neuroprotective; osteopathic; antiarthritic;
 KW antiparasitic; antihelminthic; antipsoriatic; uropathic;
 KW ophthalmological; antirheumatic; haemostatic; antibacterial; virucide;
 KW protozoacide; fungicide; gene therapy; cell proliferative disorder;
 KW arteriosclerosis; atherosclerosis; cirrhosis; hepatitis;
 KW paroxysmal nocturnal haemoglobinuria; polycythaemia vera; psoriasis;
 KW primary thrombocytopaenia; cancer; developmental disorder;
 KW renal tubular acidosis; anaemia; mental retardation;
 KW neurological disorder; Alzheimer's disease; Parkinson's disease;
 KW epilepsy; autoimmune disorder; inflammatory disorder; AIDS; allergy;
 KW asthma; autoimmune thyroiditis; contact dermatitis; Crohn's disease;
 KW diabetes mellitus; glomerulonephritis; Goodpasture's syndrome; gout;
 KW Graves' disease; Hashimoto's thyroiditis; irritable bowel syndrome;
 KW multiple sclerosis; osteoarthritis; osteoporosis; pancreatitis;
 KW Reiter's syndrome; rheumatoid arthritis; Sjogren's syndrome; uveitis;
 KW infection.

XX

OS Homo sapiens.

XX

PN WO2003068943-A2.

XX

PD 21-AUG-2003.

XX

PF 13-FEB-2003; 2003WO-US004712.

XX

PR 13-FEB-2002; 2002US-0357002P.

PR 06-MAR-2002; 2002US-0362439P.

PR 19-MAR-2002; 2002US-0366041P.

XX

PA (INCY-) INCYTE GENOMICS INC.

XX

PI Lehr-Mason PM, Kable AE, Elliott VS, Marquis JP, Baughn MR;
 PI Chawla NK, Tran UK, Jin P, Tang YT, Zebarjadian Y, Swarnakar A;

PI Hafalia AJA, Cocks BG, Warren BA, Emerling BM, Pearson CI, Chien D;
PI Peterson DP, Fu GK, Yue H, Jackson AA, Jiang X, Hawkins PR, Lal PG;
PI Khare R, Lee S, Lee SY, Richardson TW, Chang H;

XX

DR WPI; 2003-689669/65.

DR N-PSDB; ADA21192.

XX

PT New human secreted proteins and polynucleotides, useful for diagnosing,
PT treating or preventing autoimmune or inflammatory disorders (e.g. AIDS,
PT allergy, asthma or anemia), multiple sclerosis, osteoporosis, cancer or
PT hepatitis.

XX

PS Claim 1; Page 249-252; 295pp; English.

XX

CC The present sequence represents a human secreted protein (I) designated
CC SECP-46. (I) have anti-HIV, antiallergic, antiinflammatory, antianaemic,
CC antiparkinsonian, nootropic, anticonvulsant, antiarteriosclerotic,
CC antiasthmatic, immunosuppressive, antithyroid, cytostatic, hepatotropic,
CC dermatological, antidiabetic, nephrotropic, antigout, thyromimetic,
CC neuroprotective, osteopathic, antiarthritic, antiparasitic,
CC antihelminthic, antipsoriatic, uropathic, ophthalmological,
CC antirheumatic, haemostatic, antibacterial, virucide, protozoacide and
CC fungicide activities, and can be used in gene therapy. The human secreted
CC proteins (SECP), polynucleotides, agonists and antagonists of the present
CC invention are useful for diagnosing, treating or preventing disorders
CC associated with aberrant expression of SECP, particularly cell
CC proliferative disorders (e.g. arteriosclerosis, atherosclerosis,
CC cirrhosis, hepatitis, paroxysmal nocturnal haemoglobinuria, polycythaemia
CC vera, psoriasis, primary thrombocytopaenia or cancer), developmental
CC disorders (e.g. renal tubular acidosis, anaemia or mental retardation),
CC neurological disorders (e.g. Alzheimer's disease, Parkinson's disease or
CC epilepsy), autoimmune/inflammatory disorders (e.g. AIDS, allergies,
CC asthma, autoimmune thyroiditis, contact dermatitis, Crohn's disease,
CC diabetes mellitus, glomerulonephritis, Goodpasture's syndrome, gout,
CC Graves' disease, Hashimoto's thyroiditis, irritable bowel syndrome,
CC multiple sclerosis, osteoarthritis, osteoporosis, pancreatitis, Reiter's
CC syndrome, rheumatoid arthritis, Sjogren's syndrome, uveitis), or viral,
CC bacterial, fungal, parasitic, protozoan or helminthic infections. The
CC SECP and polynucleotides are also useful in assessing the effects of
CC exogenous compounds on the expression of nucleic acids secreted proteins.
CC The polynucleotides encoding SECP are useful for creating transgenic
CC animals to model human disease.

XX

SQ Sequence 1350 AA;

Query Match 37.7%; Score 1356.5; DB 6; Length 1350;
Best Local Similarity 42.6%; Pred. No. 6.1e-55;
Matches 232; Conservative 50; Mismatches 207; Indels 56; Gaps 10;

Qy 92 SQC---CP-GFYESGEMCVPHCADKCVH-GRC-IAPNTCQCEPGWGGTNCSSACDGDHWG 145
|:| || |:| | | |: | :| | | | || | |:| ||| |||
Db 706 SRCQDVCPAGWY--GPSCQTRCS--CANDGHCHPATGHCSCAPGWTGFSCQACDTGHWG 761
Qy 146 PHCTSRCQCKNG-ALCNFITGACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCQNGATCD 204
| |: | | | |: |:| | | ||: | ||| |:| | | | ||| |:| ||
Db 762 PDCSHPCNCSAGHGSCDAISGLCLCEAGYVGPRCEQQCPQGHFGPGCEQLCQCQHGAAACD 821

Qy 205 HVTGECRCPPGYTGAFCEDLCPGKHGPQCEQRCPCQNGGVCHHVTGECSCPSGWMGTVC 264
 Db 822 HVSGACTCPAGWRGTFCEHACPAFFGLDCRSACNCTAGAACDAVNGSCLCPAGRRGPRC 881

Qy 265 GQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP----- 300
 Db 882 AETCPAHTYGHNCNQACACFNGASCDPVHGGQCHCAPGWMGPSCLQECLPRDVRAGCRHSG 941

Qy 301 -----GYTGERCQDECPVGTGVLCAETCQCVNNGKCYHVSGACLC 341
 Db 942 GCLNGLCDPHTGRCLCPAGWTGDKCQSPCLRGWFGEACAQRCSCPPGAACHHVTGACRC 1001

Qy 342 EAGFAGERCEARLCPEGLYGIKCDKRCPCHELENTSCHPMSEGCACKPGWSGLYCNETCS 401
 Db 1002 PPGFTGSGCE-QACPPGSFGEDECAQMCQCPGENP-ACHPATGTCSCAAGYHGPSCQQRCP 1059

Qy 402 PGFYGEACQQICSCQNGADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKND 461
 Db 1060 PGRYGPGEQQLCGCLNNGSCDAATGACRCPTGFLGTDCNLTCQGRFGPNCTHVCGCGQG 1119

Qy 462 AVCSFVDGSCTCKAGWHGVDCSIRCPSGTWGFGCNLTQCLNNGACNTLDGTCTCAPGWR 521
 Db 1120 AACDPVTGTCLCPPGRAGVRCERGCPQNRFGVGCEHTCSCRNGGLCHASNGSCSCGLGWT 1179

Qy 522 GEKCELPCQDGTGYNCAERCDCSHADGCHPTTGHCRCPLPGWSGVHCDSVCAEGRWGPNC 581
 Db 1180 GRHCELACPPGRYGAACHLECSCHNNSTCEPATGTCTRCGPGFYQACEHPCPPGFHGAGC 1239

Qy 582 SLPCY 586
 Db 1240 QGLCW 1244

RESULT 9

ADD78227

ID ADD78227 standard; protein; 1261 AA.

XX

AC ADD78227;

XX

DT 29-JAN-2004 (first entry)

XX

DE Human CGDD-8.

XX

KW Anabolic; Hypertensive; Respiratory; Anti-HIV; Antiallergic;
 KW Neuroprotective; Nootropic; Antianemic; Antiartherosclerotic;
 KW Antiinflammatory; Ophthalmological; Muscular; Hepatotropic;
 KW Neuroprotective; Antiasthmatic; Anticonvulsant; Virucide; Antibacterial;
 KW Fungicide; Antiparasitic; Protozoacide; Antihelminthic; Cytostatic;
 KW Cerebroprotective; Antiparkinsonian; Antipsoriatic; Antigout;
 KW Antidiabetic; Antiarthritic; Antirheumatic; Osteopathic; Gene therapy;
 KW human; cell growth; cell differentiation; cell death; CGDD;
 KW cell proliferative disorder; cancer; developmental disorder;
 KW neurological disorder; autoimmune disorder; inflammatory disorder;
 KW infection; reproductive disorder.

XX

OS Homo sapiens.

XX

PN WO2003077875-A2.
 XX
 PD 25-SEP-2003.
 XX
 PF 14-MAR-2003; 2003WO-US008310.
 XX
 PR 15-MAR-2002; 2002US-0364494P.
 PR 29-MAR-2002; 2002US-0369129P.
 PR 12-APR-2002; 2002US-0372511P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Kable AE, Tran UK, Hafalia AJA, Burford N, Honchell CD;
 PI Lehr-Mason PM, Duggan BM, Ramkumar J, Griffin JA, Richardson TW;
 PI Elliott VS, Jiang X, Jackson AA, Marquis JP, Chawla NK, Khare R;
 PI Becha SD, Lee SY, Swarnakar A, Yue H, Warren BA, Baughn MR, Lal PG;
 PI Lee S, Ho A, Gandhi AR, Yao MG;
 XX
 DR WPI; 2003-779081/73.
 DR N-PSDB; ADD78266.
 XX
 PT New polypeptides and polynucleotides associated with cell growth,
 PT differentiation and death, useful for diagnosing, treating or preventing
 PT e.g. developmental, neurological, autoimmune, inflammatory or
 PT reproductive disorders.
 XX
 PS Claim 1; SEQ ID NO 8; 320pp; English.
 XX
 CC The present invention relates to novel human proteins (I; ADD78220-
 CC ADD78258) and their coding sequences (II; ADD78259-ADD78297), which are
 CC associated with cell growth, differentiation and death, referred to as
 CC CGDD-n proteins, where n is a number from 1 to 39. The CGDD proteins and
 CC their coding sequences are useful for diagnosing, treating or preventing
 CC cell proliferative disorders (e.g. cirrhosis, hepatitis,
 CC arteriosclerosis, psoriasis, primary thrombocytopenia) or cancers (e.g.
 CC adenocarcinoma, sarcoma or cancers of the bone, bone marrow, brain,
 CC breast, colon, kidney, liver, lung or uterus), developmental disorders
 CC (e.g. renal tubular acidosis, Becker muscular dystrophy, gonadal
 CC dysgenesis, hypothyroidism or seizures), neurological disorders (e.g.
 CC Pick's disease, cataract, epilepsy, ischemic cerebrovascular disease,
 CC stroke, Alzheimer's disease, Parkinson's disease or dementia),
 CC autoimmune/inflammatory disorders (e.g. AIDS, allergies, anemia, asthma,
 CC diabetes mellitus, bronchitis, osteoporosis, osteoarthritis, rheumatoid
 CC arthritis, contact dermatitis or gout), viral, bacterial, fungal,
 CC parasitic, protozoan or helminthic infections, reproductive disorders
 CC (e.g. infertility, ectopic pregnancy, premature ovarian failure, delayed
 CC puberty or prostatitis) or disorders of the placenta (e.g. preeclampsia,
 CC choriocarcinoma, placenta previa, placental or maternal floor infarction
 CC or chronic villitis).
 XX
 SQ Sequence 1261 AA;

Query Match 36.5%; Score 1313.5; DB 7; Length 1261;
 Best Local Similarity 38.8%; Pred. No. 5.5e-53;
 Matches 231; Conservative 50; Mismatches 206; Indels 109; Gaps 9;

QY 94 CCPGFYESGEMC-----VPHCADKCVHGRCIAPNT-----CQCEPGWGGTNC 135

Db	523	CDPGLY--GRFCHLTCPFWAFGPGCSEEC---QCVQPHTQSCDKRDGSCSCKAGFERGERC	577
Qy	136	SSACDGDHWGPHCTSRCQCKNGALCNPITGAC--HCAAGFRGWRCEDRCEQGTYGNDCHQ	193
Db	578	QAECELGYFGPGCWQACTCPVGVACDSVSGECGKRCPAGFQGEDCGQECVPVGTFGVNCSS	637
Qy	194	RCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPPGKHGPQCEQRCPCQNGGVCHHVTGEC	253
Db	638	SCSC--GGAPCHGVTGQCRCPPGRTGEDCEADCPEGHFGPGCEQRCQCQHGAAACDHVSGAC	696
Qy	254	SCPSGWMGTVCGQPCPEGRF-----	273
Db	697	TCPAGWRGTFCEHACPAFFGLDCRSACNCTAGAACDAVNGSCLCPAGRRGPRCAETCPA	756
Qy	274	---GKNCSQECQCHNGGTCDAAATGQCHCSPGYTGERCQDECPVGTYGVLCAETCQCVNGG	330
Db	757	HTYGHNCSQACACFNGASCDPVHGQCHCAPGWMGSPCLQACPAGLYGDNCRHSCLCQNGG	816
Qy	331	KCYHVSAGACLCEAGFAGERCEAR-----LCPEGLYG	361
Db	817	TCDPVSGHCACPEGWAGLACEKECLPRDVRAGCRHSGGCLNGGLCDPHTGRCLCPAGWTG	876
Qy	362	IKCDKRC-----PCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQ	410
Db	877	DKCQSPCLRGWFGEACAQRCSCPPGAACHHVTGACRCPPGFTGSGCEQGCPPGRYGPGE	936
Qy	411	QICSCQNGADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDVAVCSVPDGS	470
Db	937	QLCGCLNGGSCDAATGACRCPTGFLGTDCNLTCPPQGRFGPNCTHVCGCGQGAACDPVTGT	996
Qy	471	CTCKAGWHGVDCSIRCPSTGWGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPCQ	530
Db	997	CLCPPGRAGVRCERGCPQNRFGVGCEHTCSCRNGGLCHASNGSCSGLGWTGRHCELACP	1056
Qy	531	DGTYGLNCAERCDCSHADGCHPTTGHCRCLPGWSGVHCDSVCAEGRWGPNCSLPCY	586
Db	1057	PGRYGAACHLECSCHNNSTCEPATGTCTCRGPGFYGQACEHPCPPGFHGAGCQGLCW	1112

RESULT 10

ABJ37903

ID ABJ37903 standard; protein; 1403 AA.

XX

AC ABJ37903;

XX

DT 22-MAY-2003 (first entry)

XX

DE NOVX protein sequence SEQ ID No 52.

XX

KW Hepatotropic; immunosuppressive; cardiant; hypertensive; tranquilizer;

KW vulnerary; virucide; antibacterial; protozoacide; fungicide; nootropic;

KW antiparasitic; neuroprotective; cerebroprotective; antiparkinsonian;

KW anticonvulsant; antiaddictive; analgesic; dermatological; keratolytic;

KW antiseborrheic; antirheumatic; antiarthritic; antiinflammatory; anti-HIV;

KW cytostatic; antiasthmatic; antipsoriatic; hypotensive; osteopathic;

KW antiulcer; anorectic; antidiabetic; antiallergic; haemostatic;

KW neuroleptic; antidepressant; antiinfertility; NOVX; human disease;
KW NOVX-associated disorder; trauma; viral; bacterial; fungal; protozoal;
KW parasitic infection; Alzheimer's disease; stroke; forensic biology;
KW immunogen; non-human transgenic animal; gene therapy.

XX

OS Unidentified.

XX

PN WO200281517-A2.

XX

PD 17-OCT-2002.

XX

PF 22-JAN-2002; 2002WO-US002064.

XX

PR 19-JAN-2001; 2001US-0262892P.

PR 23-JAN-2001; 2001US-0263598P.

PR 24-JAN-2001; 2001US-0263799P.

PR 25-JAN-2001; 2001US-0264117P.

PR 25-JAN-2001; 2001US-0264139P.

PR 26-JAN-2001; 2001US-0264478P.

PR 30-JAN-2001; 2001US-0263351P.

PR 02-MAR-2001; 2001US-0272870P.

PR 14-MAR-2001; 2001US-0275927P.

PR 14-MAR-2001; 2001US-0275990P.

PR 15-MAR-2001; 2001US-0276449P.

PR 20-MAR-2001; 2001US-0277358P.

PR 23-MAR-2001; 2001US-0278151P.

PR 29-MAR-2001; 2001US-0279857P.

PR 20-APR-2001; 2001US-0285140P.

PR 20-APR-2001; 2001US-0285141P.

PR 30-APR-2001; 2001US-0287484P.

PR 17-MAY-2001; 2001US-0291701P.

PR 08-JUN-2001; 2001US-0296960P.

PR 10-JUL-2001; 2001US-0304353P.

PR 10-JUL-2001; 2001US-0304355P.

PR 12-JUL-2001; 2001US-0304886P.

PR 09-AUG-2001; 2001US-0311289P.

PR 13-AUG-2001; 2001US-0311975P.

PR 16-AUG-2001; 2001US-0312937P.

PR 18-OCT-2001; 2001US-0330227P.

PR 29-NOV-2001; 2001US-0334198P.

XX

PA (CURA-) CURAGEN CORP.

XX

PI Decristofaro MF, Padigar M, Miller C, Tchernev V, Zhong H;
PI Zhong M, Anderson D, Ballinger R, Gerlach V, Spytek KA, Rastelli L;
PI Kekuda R, Guo X, Zerhusen B, Andrew D, Mezes P, Patturajan M;
PI Burgess CE, Eisen A, Wolenc A, Baumgartner J, Shimkets RA, Gusev V;
PI Vernet CAM, Taupier RJ, Pena C, Shenoy S, Li L, Casman S, Boldog F;
PI Fernandes E, Smithson G, Malyankar U, Taillon B, Liu X;

XX

DR WPI; 2003-058504/05.

DR N-PSDB; ABT33368.

XX

PT New polypeptides, designated as NOVX, useful for diagnosing and treating
PT infections, neurological diseases, cancer, allergy, and bone,
PT immunological, skin, renal, brain, muscle and autoimmune disorders.

XX

PS Claim 1; Page 133; 672pp; English.

XX

CC The invention relates to a novel isolated polypeptide, designated NOVX
CC (NOV1 - 33), consisting of a mature form of one of 61 sequences, given in
CC the specification, or its variant, where amino acid residue(s) in the
CC variant differ from the mature form, provided that the variant differs in
CC not more than 15 % of the amino acids from the sequence of the mature
CC form. The NOVX polypeptides, nucleic acids encoding the polypeptides, and
CC an antibody to the polypeptides, are useful for treating or preventing a
CC NOVX-associated disorder in humans and for treating a syndrome associated
CC with a human disease (NOVX-associated disorder). NOVX polypeptides and
CC the encoding nucleic acids, are useful for determining the presence of or
CC predisposition to a disease associated with altered levels of NOVX
CC polypeptide and polynucleotide, by measuring the level of polypeptide
CC expression or the amount of nucleic acid from a mammal and comparing it
CC with another mammal not having or not predisposed to the disease. NOVX
CC polypeptide is also useful for identifying an agent that binds to NOVX
CC and a cell expressing NOVX is useful for identifying an agent that
CC modulates the expression or activity of NOVX. The antibodies and a
CC polypeptide having 95 % sequence identity to NOVX polypeptide are useful
CC for treating a pathological state in a mammal. The antibodies are also
CC useful for determining the presence or amount of NOVX in a sample. NOVX
CC polypeptides, polynucleotides and antibodies specific for the
CC polypeptides are useful for treating or preventing disorders or syndromes
CC including trauma, viral, bacterial, fungal, protozoal, and parasitic
CC infections. They can also treat disorders such as e.g., Alzheimer's
CC disease or a stroke. The NOVX encoding nucleic acids are useful for
CC expressing the NOVX proteins, to detect NOVX mRNA, or a genetic lesion in
CC a NOVX gene and to modulate NOVX activity. NOVX sequences are also useful
CC for identifying a cell or tissue type in a biological sample, to amplify
CC DNA sequences from very small biological samples such as tissues e.g.
CC hair or skin or body fluids in forensic biology and as primers and probes
CC for use in identifying and/or cloning NOVX homologues in other cell
CC types. The NOVX proteins are useful as an immunogen to generate
CC antibodies which are useful for diagnostically monitoring protein levels
CC and modulating NOVX activity. Cells comprising NOVX nucleic acids are
CC useful for producing non-human transgenic animals which are useful for
CC studying the function and/or activity of NOVX protein and for identifying
CC and/or evaluating modulators of NOVX protein activity. The NOVX nucleic
CC acids can be used in gene therapy. This sequence represents a NOVX
CC protein of the invention

XX

SQ Sequence 1403 AA;

Query Match 36.4%; Score 1310.5; DB 6; Length 1403;
Best Local Similarity 41.8%; Pred. No. 8.2e-53;
Matches 228; Conservative 45; Mismatches 217; Indels 55; Gaps 10;

QY 90 RKSQCCPGFYESGEMCVPHCADKCVH-GRCIA-PNTCQCEPGWGGTNCSSACDGDHWGPH 147
| | :| | | : | : | | | | | :| | | |
Db 814 RCQDCEAGWY--GPSCQTMCS--CANDGHCHQDTGHCSCAPGWTGFSCQRACDTGHWGPD 869
QY 148 CTSRCQCKNG-ALCNPITGACHCAAGFRGWRCE-DRCEQGTYGNDCHQRCQCQNGATCDH 205
| : | | | : | : | | | : | | | :| | | | :| | |
Db 870 CSHPCNCSAGHGSCDAISGLCLCEAGYVGPCEQSECPQGHFGPGCEQRCQCQHGAAACDH 929
QY 206 VTGECRCPPGYTGAFCEDLCPGKHGPQCEQRCPCQNGGVCHHVTGECSCPSGWMGTVCV 265

Db	930	VSGACTCPAGWRGTFCEHACPAGFFGLDCRSACNCTAGAACDAVNGSCLCPAGRRGPRCA	989
Qy	266	QPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSPGYTGERCQDECPVGTYGVLCAETCQ	325
Db	990	ETCPAHTYGHNCSQACACFENGASCDPVHVGQCHCAPGWMGPSCLQACPAGLYGDNCRHSCL	1049
Qy	326	CVNGGKCYHVSAGACLCEAGFAGERCEAR-----LCP	356
Db	1050	CQNGGTCDPVS GHACPEGWAGLACEVECLPRDVRAGCRHSGGCLNGGLCDPHTGRCLCP	1109
Qy	357	-----EGLYGIKCDKRCPCHELENTHSCHPMSEACCKPGWSGLYCNETCS	401
Db	1110	AGWTGDKCQSPAACAKGTFGPHCEGRACACRWGG--PCHLATGACLCPGWRGPHLSAACL	1167
Qy	402	PGFYGEACQQICSCQNGADCDVSTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKND	461
Db	1168	RGWFGEACAQRCS CPPGAACHVVTGACRCPPGFTGSGCEQACPPGSFGECAQMCQCPGE	1227
Qy	462	-AVCSPVDGSCCTCKAGWHGVDCSIRCPSGTWGFGCNLTCQCLNGGACNTLDGTCTCAPGW	520
Db	1228	NPACHPATGTCS CAAGYHGPSCQQRCPGGRYGPGEQLCGCLNGGSCDAATGACRCPTGF	1287
Qy	521	RGEKCELPQDGTGYGLNCAERCDCSHADGCHPTTGHCRCPLPGWSGVHCDSVCAEGRWGP	580
Db	1288	LGTDCNLTC PQGRFGPNCTHVC GCGQGAACDPVTGTCLCPPGRAGVRCERGC PQNRFGVG	1347
Qy	581	CSLPC	585
Db	1348	CEHTC	1352

RESULT 11

ABG08033

ID ABG08033 standard; protein; 878 AA.

XX

AC ABG08033;

XX

DT 13-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #8024.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US008631.

XX

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX

PA (HYSE-) HYSEQ INC.

Qy 576 RWGPNCSLPCY 586
 |||||||
Db 231 RWGPNCSLPCY 241

RESULT 12

ABJ37904

ID ABJ37904 standard; protein; 1577 AA.

XX

AC ABJ37904;

XX

DT 22-MAY-2003 (first entry)

XX

DE NOVX protein sequence SEQ ID No 54.

XX

KW Hepatotropic; immunosuppressive; cardiant; hypertensive; tranquilizer;
KW vulnerary; virucide; antibacterial; protozoacide; fungicide; nootropic;
KW antiparasitic; neuroprotective; cerebroprotective; antiparkinsonian;
KW anticonvulsant; antiaddictive; analgesic; dermatological; keratolytic;
KW antiseborrheic; antirheumatic; antiarthritic; antiinflammatory; anti-HIV;
KW cytostatic; antiasthmatic; antipsoriatic; hypotensive; osteopathic;
KW antiulcer; anorectic; antidiabetic; antiallergic; haemostatic;
KW neuroleptic; antidepressant; antiinfertility; NOVX; human disease;
KW NOVX-associated disorder; trauma; viral; bacterial; fungal; protozoal;
KW parasitic infection; Alzheimer's disease; stroke; forensic biology;
KW immunogen; non-human transgenic animal; gene therapy.

XX

OS Unidentified.

XX

PN WO200281517-A2.

XX

PD 17-OCT-2002.

XX

PF 22-JAN-2002; 2002WO-US002064.

XX

PR 19-JAN-2001; 2001US-0262892P.

PR 23-JAN-2001; 2001US-0263598P.

PR 24-JAN-2001; 2001US-0263799P.

PR 25-JAN-2001; 2001US-0264117P.

PR 25-JAN-2001; 2001US-0264139P.

PR 26-JAN-2001; 2001US-0264478P.

PR 30-JAN-2001; 2001US-0263351P.

PR 02-MAR-2001; 2001US-0272870P.

PR 14-MAR-2001; 2001US-0275927P.

PR 14-MAR-2001; 2001US-0275990P.

PR 15-MAR-2001; 2001US-0276449P.

PR 20-MAR-2001; 2001US-0277358P.

PR 23-MAR-2001; 2001US-0278151P.

PR 29-MAR-2001; 2001US-0279857P.

PR 20-APR-2001; 2001US-0285140P.

PR 20-APR-2001; 2001US-0285141P.

PR 30-APR-2001; 2001US-0287484P.

PR 17-MAY-2001; 2001US-0291701P.

PR 08-JUN-2001; 2001US-0296960P.

PR 10-JUL-2001; 2001US-0304353P.

PR 10-JUL-2001; 2001US-0304355P.

PR 12-JUL-2001; 2001US-0304886P.

PR 09-AUG-2001; 2001US-0311289P.
PR 13-AUG-2001; 2001US-0311975P.
PR 16-AUG-2001; 2001US-0312937P.
PR 18-OCT-2001; 2001US-0330227P.
PR 29-NOV-2001; 2001US-0334198P.

XX

PA (CURA-) CURAGEN CORP.

XX

PI Decristofaro MF, Padigaru M, Miller C, Tchernev V, Zhong H;
PI Zhong M, Anderson D, Ballinger R, Gerlach V, Spytek KA, Rastelli L;
PI Kekuda R, Guo X, Zerhusen B, Andrew D, Mezes P, Patturajan M;
PI Burgess CE, Eisen A, Wolenc A, Baumgartner J, Shimkets RA, Gusev V;
PI Vernet CAM, Taupier RJ, Pena C, Shenoy S, Li L, Casman S, Boldog F;
PI Fernandes E, Smithson G, Malyankar U, Taillon B, Liu X;

XX

DR WPI; 2003-058504/05.

DR N-PSDB; ABT33369.

XX

PT New polypeptides, designated as NOVX, useful for diagnosing and treating
PT infections, neurological diseases, cancer, allergy, and bone,
PT immunological, skin, renal, brain, muscle and autoimmune disorders.

XX

PS Claim 1; Page 135-136; 672pp; English.

XX

CC The invention relates to a novel isolated polypeptide, designated NOVX
CC (NOV1 - 33), consisting of a mature form of one of 61 sequences, given in
CC the specification, or its variant, where amino acid residue(s) in the
CC variant differ from the mature form, provided that the variant differs in
CC not more than 15 % of the amino acids from the sequence of the mature
CC form. The NOVX polypeptides, nucleic acids encoding the polypeptides, and
CC an antibody to the polypeptides, are useful for treating or preventing a
CC NOVX-associated disorder in humans and for treating a syndrome associated
CC with a human disease (NOVX-associated disorder). NOVX polypeptides and
CC the encoding nucleic acids, are useful for determining the presence of or
CC predisposition to a disease associated with altered levels of NOVX
CC polypeptide and polynucleotide, by measuring the level of polypeptide
CC expression or the amount of nucleic acid from a mammal and comparing it
CC with another mammal not having or not predisposed to the disease. NOVX
CC polypeptide is also useful for identifying an agent that binds to NOVX
CC and a cell expressing NOVX is useful for identifying an agent that
CC modulates the expression or activity of NOVX. The antibodies and a
CC polypeptide having 95 % sequence identity to NOVX polypeptide are useful
CC for treating a pathological state in a mammal. The antibodies are also
CC useful for determining the presence or amount of NOVX in a sample. NOVX
CC polypeptides, polynucleotides and antibodies specific for the
CC polypeptides are useful for treating or preventing disorders or syndromes
CC including trauma, viral, bacterial, fungal, protozoal, and parasitic
CC infections. They can also treat disorders such as e.g., Alzheimer's
CC disease or a stroke. The NOVX encoding nucleic acids are useful for
CC expressing the NOVX proteins, to detect NOVX mRNA, or a genetic lesion in
CC a NOVX gene and to modulate NOVX activity. NOVX sequences are also useful
CC for identifying a cell or tissue type in a biological sample, to amplify
CC DNA sequences from very small biological samples such as tissues e.g.
CC hair or skin or body fluids in forensic biology and as primers and probes
CC for use in identifying and/or cloning NOVX homologues in other cell
CC types. The NOVX proteins are useful as an immunogen to generate
CC antibodies which are useful for diagnostically monitoring protein levels

RESULT 13

ABJ37900

ID ABJ37900 standard; protein; 1398 AA.

XX

AC ABJ37900;

XX

DT 22-MAY-2003 (first entry)

XX

DE NOVX protein sequence SEQ ID No 46.

XX

KW Hepatotropic; immunosuppressive; cardiant; hypertensive; tranquilizer;

KW vulnerary; virucide; antibacterial; protozoacide; fungicide; nootropic;

KW antiparasitic; neuroprotective; cerebroprotective; antiparkinsonian;

KW anticonvulsant; antiaddictive; analgesic; dermatological; keratolytic;

KW antiseborrheic; antirheumatic; antiarthritic; antiinflammatory; anti-HIV;

KW cytostatic; antiasthmatic; antipsoriatic; hypotensive; osteopathic;

KW antiulcer; anorectic; antidiabetic; antiallergic; haemostatic;

KW neuroleptic; antidepressant; antiinfertility; NOVX; human disease;

KW NOVX-associated disorder; trauma; viral; bacterial; fungal; protozoal;

KW parasitic infection; Alzheimer's disease; stroke; forensic biology;

KW immunogen; non-human transgenic animal; gene therapy.

XX

OS Unidentified.

XX

PN WO200281517-A2.

XX

PD 17-OCT-2002.

XX

PF 22-JAN-2002; 2002WO-US002064.

XX

PR 19-JAN-2001; 2001US-0262892P.

PR 23-JAN-2001; 2001US-0263598P.

PR 24-JAN-2001; 2001US-0263799P.

PR 25-JAN-2001; 2001US-0264117P.

PR 25-JAN-2001; 2001US-0264139P.

PR 26-JAN-2001; 2001US-0264478P.

PR 30-JAN-2001; 2001US-0263351P.

PR 02-MAR-2001; 2001US-0272870P.

PR 14-MAR-2001; 2001US-0275927P.

PR 14-MAR-2001; 2001US-0275990P.

PR 15-MAR-2001; 2001US-0276449P.

PR 20-MAR-2001; 2001US-0277358P.

PR 23-MAR-2001; 2001US-0278151P.

PR 29-MAR-2001; 2001US-0279857P.

PR 20-APR-2001; 2001US-0285140P.

PR 20-APR-2001; 2001US-0285141P.

PR 30-APR-2001; 2001US-0287484P.

PR 17-MAY-2001; 2001US-0291701P.

PR 08-JUN-2001; 2001US-0296960P.

PR 10-JUL-2001; 2001US-0304353P.

PR 10-JUL-2001; 2001US-0304355P.

PR 12-JUL-2001; 2001US-0304886P.

PR 09-AUG-2001; 2001US-0311289P.

PR 13-AUG-2001; 2001US-0311975P.

PR 16-AUG-2001; 2001US-0312937P.

PR 18-OCT-2001; 2001US-0330227P.

PR 29-NOV-2001; 2001US-0334198P.

XX

PA (CURA-) CURAGEN CORP.

XX

PI Decristofaro MF, Padigar M, Miller C, Tchernev V, Zhong H;
PI Zhong M, Anderson D, Ballinger R, Gerlach V, Spytek KA, Rastelli L;
PI Kekuda R, Guo X, Zerhusen B, Andrew D, Mezes P, Patturajan M;
PI Burgess CE, Eisen A, Wolenc A, Baumgartner J, Shimkets RA, Gusev V;
PI Vernet CAM, Taupier RJ, Pena C, Shenoy S, Li L, Casman S, Boldog F;
PI Fernandes E, Smithson G, Malyankar U, Taillon B, Liu X;

XX

DR WPI; 2003-058504/05.

DR N-PSDB; ABT33365.

XX

PT New polypeptides, designated as NOVX, useful for diagnosing and treating
PT infections, neurological diseases, cancer, allergy, and bone,
PT immunological, skin, renal, brain, muscle and autoimmune disorders.

XX

PS Claim 1; Page 127; 672pp; English.

XX

CC The invention relates to a novel isolated polypeptide, designated NOVX
CC (NOV1 - 33), consisting of a mature form of one of 61 sequences, given in
CC the specification, or its variant, where amino acid residue(s) in the
CC variant differ from the mature form, provided that the variant differs in
CC not more than 15 % of the amino acids from the sequence of the mature
CC form. The NOVX polypeptides, nucleic acids encoding the polypeptides, and
CC an antibody to the polypeptides, are useful for treating or preventing a
CC NOVX-associated disorder in humans and for treating a syndrome associated
CC with a human disease (NOVX-associated disorder). NOVX polypeptides and
CC the encoding nucleic acids, are useful for determining the presence of or
CC predisposition to a disease associated with altered levels of NOVX
CC polypeptide and polynucleotide, by measuring the level of polypeptide
CC expression or the amount of nucleic acid from a mammal and comparing it
CC with another mammal not having or not predisposed to the disease. NOVX
CC polypeptide is also useful for identifying an agent that binds to NOVX
CC and a cell expressing NOVX is useful for identifying an agent that
CC modulates the expression or activity of NOVX. The antibodies and a
CC polypeptide having 95 % sequence identity to NOVX polypeptide are useful
CC for treating a pathological state in a mammal. The antibodies are also
CC useful for determining the presence or amount of NOVX in a sample. NOVX
CC polypeptides, polynucleotides and antibodies specific for the
CC polypeptides are useful for treating or preventing disorders or syndromes
CC including trauma, viral, bacterial, fungal, protozoal, and parasitic
CC infections. They can also treat disorders such as e.g., Alzheimer's
CC disease or a stroke. The NOVX encoding nucleic acids are useful for
CC expressing the NOVX proteins, to detect NOVX mRNA, or a genetic lesion in
CC a NOVX gene and to modulate NOVX activity. NOVX sequences are also useful
CC for identifying a cell or tissue type in a biological sample, to amplify
CC DNA sequences from very small biological samples such as tissues e.g.
CC hair or skin or body fluids in forensic biology and as primers and probes
CC for use in identifying and/or cloning NOVX homologues in other cell
CC types. The NOVX proteins are useful as an immunogen to generate
CC antibodies which are useful for diagnostically monitoring protein levels
CC and modulating NOVX activity. Cells comprising NOVX nucleic acids are
CC useful for producing non-human transgenic animals which are useful for
CC studying the function and/or activity of NOVX protein and for identifying
CC and/or evaluating modulators of NOVX protein activity. The NOVX nucleic

CC acids can be used in gene therapy. This sequence represents a NOVX
CC protein of the invention
XX
SQ Sequence 1398 AA;

Query Match 36.1%; Score 1300; DB 6; Length 1398;
Best Local Similarity 41.8%; Pred. No. 2.5e-52;
Matches 228; Conservative 45; Mismatches 217; Indels 56; Gaps 11;

```
Qy      90 RKSQCCPGFYESGEMCVPHCADKCVH-GRCIA-PNTCQCEPGWGGTNCSSACDGDHWGPH 147
      | | | : | | | : | : | | | | | | | : | | | | | |
Db      808 RCQDCEAGWY--GPSCQTMCS--CANDGHCHQDTGHCSCAPGWTGFSCQRACDTGHWGPD 863

Qy      148 CTSRCQCKNG-ALCNPITGACHCAAGFRGWRCE-DRCEQGTYGNDCHQRCQCQNGATCDH 205
      | : | | | | : | : | | | | : | | | | | : | | | |
Db      864 CSHPCNCSAGHGSCDAISGLCLCEAGYVGPRCEQSECPQGHFGPGCEQRCQCQHGAAACDH 923

Qy      206 VTGECRCPPGYTGAFCEDLCPGKGHPQCEQRCPCQNGGVCHHVTGECSCPSGWMGTVCV 265
      | : | | | | : | | | | | | | | | | | | | | : | | |
Db      924 VSGACTCPAGWRGTFCFHACPAFFGLDCRSACNCTAGAACDAVNGSCLCPAGRRGPRCA 983

Qy      266 Q-PCPEGRFGKNCSSQECQCHNGGTCDAAATGQCHCSPGYTGERCQDECPVGTYGVLCAETC 324
      : | | : | | | | | | | : | | | | | : | | | | | : |
Db      984 ESACPAHTYGHNCSSQACACFNGASCDPVHGGQCHCAPGWMGPSCLQACPAGLYGDNCRHSC 1043

Qy      325 QCVNNGGKCYHVSGACLCEAGFAGERCEAR-----LC 355
      | | | | | | | | | : | | | | | | | | | | |
Db      1044 LCQNGGTCDPVSGHCACPEGWAGLACEVECLPRDVRAGCRHSGGCLNGLCDPHTGRCLC 1103

Qy      356 P-----EGLYGIKCDKRCPCHELENTHSCHPMSSGECACKPGWSGLYCNETC 400
      | : | : | : | | | | | | | | | | | | | : : |
Db      1104 PAGWTGDKCQSPAACAKGTFFGPHCEGRACRWGG--PCHLATGACLCPPGWRGPHLSAAC 1161

Qy      401 SPGFYGEACQQICSCQNGADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKN 460
      | : | | | | | | | | | | | | | | | | | | : : | : |
Db      1162 LRGWTFGEACAQRCSCPPGAACHHVTGACRCPPGFTGSGCEQACPPGSFGEDCAQMCQCPG 1221

Qy      461 D-AVCSPVDGSCCTCKAGWHGVDCSIRCPSGTWGFGCNLTQCCLNGGACNTLDGTCTCAPG 519
      : | | | : | : | | : | | | | | | | | | | : | | |
Db      1222 ENPACHPATGTCTSCAAGYHGPSCQQRCPPGRYGPGEQLCGCLNNGSCDAATGACRCPTG 1281

Qy      520 WRGEKCELPCQDGTGTYGLNCAERDCSHADGCHPTTGHCRCLPGWSGVHCDSVCAEGRWGP 579
      : | | | | | : | | | | | | | | | | | : | | : | : |
Db      1282 FLGTDCNLTCPPQGRFGPNCTHVCGCGQGAACDPVTGTCLCPPGRAGVRCERGCPQNRFGV 1341

Qy      580 NCSLPC 585
      | |
Db      1342 GCEHTC 1347
```

RESULT 14

ABJ37899

ID ABJ37899 standard; protein; 1404 AA.

XX

AC ABJ37899;

XX

DT 22-MAY-2003 (first entry)

XX
 DE NOVX protein sequence SEQ ID No 44.
 XX
 KW Hepatotropic; immunosuppressive; cardiant; hypertensive; tranquilizer;
 KW vulnerary; virucide; antibacterial; protozoacide; fungicide; nootropic;
 KW antiparasitic; neuroprotective; cerebroprotective; antiparkinsonian;
 KW anticonvulsant; antiaddictive; analgesic; dermatological; keratolytic;
 KW antiseborrheic; antirheumatic; antiarthritic; antiinflammatory; anti-HIV;
 KW cytostatic; antiasthmatic; antipsoriatic; hypotensive; osteopathic;
 KW antiulcer; anorectic; antidiabetic; antiallergic; haemostatic;
 KW neuroleptic; antidepressant; antiinfertility; NOVX; human disease;
 KW NOVX-associated disorder; trauma; viral; bacterial; fungal; protozoal;
 KW parasitic infection; Alzheimer's disease; stroke; forensic biology;
 KW immunogen; non-human transgenic animal; gene therapy.
 XX
 OS Unidentified.
 XX
 PN WO200281517-A2.
 XX
 PD 17-OCT-2002.
 XX
 PF 22-JAN-2002; 2002WO-US002064.
 XX
 PR 19-JAN-2001; 2001US-0262892P.
 PR 23-JAN-2001; 2001US-0263598P.
 PR 24-JAN-2001; 2001US-0263799P.
 PR 25-JAN-2001; 2001US-0264117P.
 PR 25-JAN-2001; 2001US-0264139P.
 PR 26-JAN-2001; 2001US-0264478P.
 PR 30-JAN-2001; 2001US-0263351P.
 PR 02-MAR-2001; 2001US-0272870P.
 PR 14-MAR-2001; 2001US-0275927P.
 PR 14-MAR-2001; 2001US-0275990P.
 PR 15-MAR-2001; 2001US-0276449P.
 PR 20-MAR-2001; 2001US-0277358P.
 PR 23-MAR-2001; 2001US-0278151P.
 PR 29-MAR-2001; 2001US-0279857P.
 PR 20-APR-2001; 2001US-0285140P.
 PR 20-APR-2001; 2001US-0285141P.
 PR 30-APR-2001; 2001US-0287484P.
 PR 17-MAY-2001; 2001US-0291701P.
 PR 08-JUN-2001; 2001US-0296960P.
 PR 10-JUL-2001; 2001US-0304353P.
 PR 10-JUL-2001; 2001US-0304355P.
 PR 12-JUL-2001; 2001US-0304886P.
 PR 09-AUG-2001; 2001US-0311289P.
 PR 13-AUG-2001; 2001US-0311975P.
 PR 16-AUG-2001; 2001US-0312937P.
 PR 18-OCT-2001; 2001US-0330227P.
 PR 29-NOV-2001; 2001US-0334198P.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Decristofaro MF, Padigaru M, Miller C, Tchernev V, Zhong H;
 PI Zhong M, Anderson D, Ballinger R, Gerlach V, Spytek KA, Rastelli L;
 PI Kekuda R, Guo X, Zerhusen B, Andrew D, Mezes P, Patturajan M;
 PI Burgess CE, Eisen A, Wolenc A, Baumgartner J, Shimkets RA, Gusev V;

PI Vernet CAM, Taupier RJ, Pena C, Shenoy S, Li L, Casman S, Boldog F;
PI Fernandes E, Smithson G, Malyankar U, Taillon B, Liu X;
XX
DR WPI; 2003-058504/05.
DR N-PSDB; ABT33364.

XX
PT New polypeptides, designated as NOVX, useful for diagnosing and treating
PT infections, neurological diseases, cancer, allergy, and bone,
PT immunological, skin, renal, brain, muscle and autoimmune disorders.

XX
PS Claim 1; Page 124; 672pp; English.

XX
CC The invention relates to a novel isolated polypeptide, designated NOVX
CC (NOV1 - 33), consisting of a mature form of one of 61 sequences, given in
CC the specification, or its variant, where amino acid residue(s) in the
CC variant differ from the mature form, provided that the variant differs in
CC not more than 15 % of the amino acids from the sequence of the mature
CC form. The NOVX polypeptides, nucleic acids encoding the polypeptides, and
CC an antibody to the polypeptides, are useful for treating or preventing a
CC NOVX-associated disorder in humans and for treating a syndrome associated
CC with a human disease (NOVX-associated disorder). NOVX polypeptides and
CC the encoding nucleic acids, are useful for determining the presence of or
CC predisposition to a disease associated with altered levels of NOVX
CC polypeptide and polynucleotide, by measuring the level of polypeptide
CC expression or the amount of nucleic acid from a mammal and comparing it
CC with another mammal not having or not predisposed to the disease. NOVX
CC polypeptide is also useful for identifying an agent that binds to NOVX
CC and a cell expressing NOVX is useful for identifying an agent that
CC modulates the expression or activity of NOVX. The antibodies and a
CC polypeptide having 95 % sequence identity to NOVX polypeptide are useful
CC for treating a pathological state in a mammal. The antibodies are also
CC useful for determining the presence or amount of NOVX in a sample. NOVX
CC polypeptides, polynucleotides and antibodies specific for the
CC polypeptides are useful for treating or preventing disorders or syndromes
CC including trauma, viral, bacterial, fungal, protozoal, and parasitic
CC infections. They can also treat disorders such as e.g., Alzheimer's
CC disease or a stroke. The NOVX encoding nucleic acids are useful for
CC expressing the NOVX proteins, to detect NOVX mRNA, or a genetic lesion in
CC a NOVX gene and to modulate NOVX activity. NOVX sequences are also useful
CC for identifying a cell or tissue type in a biological sample, to amplify
CC DNA sequences from very small biological samples such as tissues e.g.
CC hair or skin or body fluids in forensic biology and as primers and probes
CC for use in identifying and/or cloning NOVX homologues in other cell
CC types. The NOVX proteins are useful as an immunogen to generate
CC antibodies which are useful for diagnostically monitoring protein levels
CC and modulating NOVX activity. Cells comprising NOVX nucleic acids are
CC useful for producing non-human transgenic animals which are useful for
CC studying the function and/or activity of NOVX protein and for identifying
CC and/or evaluating modulators of NOVX protein activity. The NOVX nucleic
CC acids can be used in gene therapy. This sequence represents a NOVX
CC protein of the invention

XX
SQ Sequence 1404 AA;

Query Match 36.1%; Score 1300; DB 6; Length 1404;
Best Local Similarity 41.8%; Pred. No. 2.5e-52;
Matches 228; Conservative 45; Mismatches 217; Indels 56; Gaps 11;

Qy 90 RKSQCCPGFYESGEMCVPHCADKCVH-GRCIA-PNTCQCEPGWGGTNCSSACDGDHWGPH 147
| | | : | | | : | : | | | | | : | | | | |
Db 814 RCQDCEAGWY--GPSCQTMCS--CANDGHCHQDTGHCSCAPGWTGFSCQACDTGHWGPD 869
Qy 148 CTSRCQCKNG-ALCNPITGACHCAAGFRGWRCE-DRCEQGTYGNDCHQRCQCNATCDH 205
| : | | | | : | : | | | | : | | | | : | | | |
Db 870 CSHPCNCSAGHGSCDAISGLCLCEAGYVGPRCEQSECPQGHFGPGCEQRCQCNHGAACDH 929
Qy 206 VTGECRCPPGYTGAFCECLCPPGKHGPQCEQRCPCQNGGVCHHVTGECSCPSGWMGTVC 265
| : | | | | : | | | | | | | | | | | | : | | |
Db 930 VSGACTCPAGWRGTFCHEACPAFFGLDCRSACNCTAGAACDAVNGSCLCPAGRRGPRCA 989
Qy 266 Q-PCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSPGYTGERCQDECPVGTYGVLCAETC 324
: | | : | | | | | | : | | | | | : | | | | : |
Db 990 ESACPAHTYGHNCQACACFNGASCDPVHVGQCHCAPGWMGPSCQLQACPAAGLYGDNCRHSC 1049
Qy 325 QCVNGGKCYHVSAGCLCEAGFAGERCEAR-----LC 355
| | | | | | | : | | | | |
Db 1050 LCQNGGTCDPVSGHCACPEGWAGLACEVECLPRDVRAGCRHSGGCLNGGLCDPHTGRCLC 1109
Qy 356 P-----EGLYGIKCDKRCPCHELENTHSCHPMSEACCKPGWSGLYCNETC 400
| : | : | : | | | | : | | | | | : : |
Db 1110 PAGWTGDKCQSPAACAKGTFGPHCEGRACACRWGG--PCHLATGACLCPPGWRGPHLSAAC 1167
Qy 401 SPGFYGEACQQICSCQNGADCDSTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKN 460
| : | | | | | | | | | | | | | | : : | : | |
Db 1168 LRGWFGCAQRCSCPPGAACHHVTGACRCPPGFTGSGCEQACPPGSFGEDCAQMCQCPG 1227
Qy 461 D-AVCSPVDGSCCTCKAGWHGVDCSIRCPSTGWFSGCNLTCQCLNGGACNTLDGTCTCAPG 519
: | | | : | : | | | | | : | | | | : | | | |
Db 1228 ENPACHPATGTCSAAGYHGPSCQQRCPGGRYGPGEQLCGCLNGGSCDAATGACRCPTG 1287
Qy 520 WRGEKCELPQDGTGTYGLNCAERCDCSHADGCHPTTGHCRCCLPGWSGVHCDVCAEGRWGP 579
: | | | | : | | | | | | | | : | : | : | : |
Db 1288 FLGTDCNLTCPPQGRFGPNCTHVCGCGQGAACDPVTGTCLCPPGRAGVRCERGCPQNRFGV 1347
Qy 580 NCSLPC 585
| |
Db 1348 GCEHTC 1353

RESULT 15

ABJ37901

ID ABJ37901 standard; protein; 1450 AA.

XX

AC ABJ37901;

XX

DT 22-MAY-2003 (first entry)

XX

DE NOVX protein sequence SEQ ID No 48.

XX

KW Hepatotropic; immunosuppressive; cardiant; hypertensive; tranquilizer;
KW vulnerary; virucide; antibacterial; protozoacide; fungicide; nootropic;
KW antiparasitic; neuroprotective; cerebroprotective; antiparkinsonian;
KW anticonvulsant; antiaddictive; analgesic; dermatological; keratolytic;
KW antiseborrheic; antirheumatic; antiarthritic; antiinflammatory; anti-HIV;

KW cytostatic; antiasthmatic; antipsoriatic; hypotensive; osteopathic;
 KW antiulcer; anorectic; antidiabetic; antiallergic; haemostatic;
 KW neuroleptic; antidepressant; antiinfertility; NOVX; human disease;
 KW NOVX-associated disorder; trauma; viral; bacterial; fungal; protozoal;
 KW parasitic infection; Alzheimer's disease; stroke; forensic biology;
 KW immunogen; non-human transgenic animal; gene therapy.
 XX
 OS Unidentified.
 XX
 PN WO200281517-A2.
 XX
 PD 17-OCT-2002.
 XX
 PF 22-JAN-2002; 2002WO-US002064.
 XX
 PR 19-JAN-2001; 2001US-0262892P.
 PR 23-JAN-2001; 2001US-0263598P.
 PR 24-JAN-2001; 2001US-0263799P.
 PR 25-JAN-2001; 2001US-0264117P.
 PR 25-JAN-2001; 2001US-0264139P.
 PR 26-JAN-2001; 2001US-0264478P.
 PR 30-JAN-2001; 2001US-0263351P.
 PR 02-MAR-2001; 2001US-0272870P.
 PR 14-MAR-2001; 2001US-0275927P.
 PR 14-MAR-2001; 2001US-0275990P.
 PR 15-MAR-2001; 2001US-0276449P.
 PR 20-MAR-2001; 2001US-0277358P.
 PR 23-MAR-2001; 2001US-0278151P.
 PR 29-MAR-2001; 2001US-0279857P.
 PR 20-APR-2001; 2001US-0285140P.
 PR 20-APR-2001; 2001US-0285141P.
 PR 30-APR-2001; 2001US-0287484P.
 PR 17-MAY-2001; 2001US-0291701P.
 PR 08-JUN-2001; 2001US-0296960P.
 PR 10-JUL-2001; 2001US-0304353P.
 PR 10-JUL-2001; 2001US-0304355P.
 PR 12-JUL-2001; 2001US-0304886P.
 PR 09-AUG-2001; 2001US-0311289P.
 PR 13-AUG-2001; 2001US-0311975P.
 PR 16-AUG-2001; 2001US-0312937P.
 PR 18-OCT-2001; 2001US-0330227P.
 PR 29-NOV-2001; 2001US-0334198P.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Decristofaro MF, Padigar M, Miller C, Tchernev V, Zhong H;
 PI Zhong M, Anderson D, Ballinger R, Gerlach V, Spytek KA, Rastelli L;
 PI Kekuda R, Guo X, Zerhusen B, Andrew D, Mezes P, Patturajan M;
 PI Burgess CE, Eisen A, Wolenc A, Baumgartner J, Shimkets RA, Gusev V;
 PI Vernet CAM, Taupier RJ, Pena C, Shenoy S, Li L, Casman S, Boldog F;
 PI Fernandes E, Smithson G, Malyankar U, Taillon B, Liu X;
 XX
 DR WPI; 2003-058504/05.
 DR N-PSDB; ABT33366.
 XX
 PT New polypeptides, designated as NOVX, useful for diagnosing and treating
 PT infections, neurological diseases, cancer, allergy, and bone,

PT immunological, skin, renal, brain, muscle and autoimmune disorders.

XX

PS Claim 1; Page 129-130; 672pp; English.

XX

CC The invention relates to a novel isolated polypeptide, designated NOVX
CC (NOV1 - 33), consisting of a mature form of one of 61 sequences, given in
CC the specification, or its variant, where amino acid residue(s) in the
CC variant differ from the mature form, provided that the variant differs in
CC not more than 15 % of the amino acids from the sequence of the mature
CC form. The NOVX polypeptides, nucleic acids encoding the polypeptides, and
CC an antibody to the polypeptides, are useful for treating or preventing a
CC NOVX-associated disorder in humans and for treating a syndrome associated
CC with a human disease (NOVX-associated disorder). NOVX polypeptides and
CC the encoding nucleic acids, are useful for determining the presence of or
CC predisposition to a disease associated with altered levels of NOVX
CC polypeptide and polynucleotide, by measuring the level of polypeptide
CC expression or the amount of nucleic acid from a mammal and comparing it
CC with another mammal not having or not predisposed to the disease. NOVX
CC polypeptide is also useful for identifying an agent that binds to NOVX
CC and a cell expressing NOVX is useful for identifying an agent that
CC modulates the expression or activity of NOVX. The antibodies and a
CC polypeptide having 95 % sequence identity to NOVX polypeptide are useful
CC for treating a pathological state in a mammal. The antibodies are also
CC useful for determining the presence or amount of NOVX in a sample. NOVX
CC polypeptides, polynucleotides and antibodies specific for the
CC polypeptides are useful for treating or preventing disorders or syndromes
CC including trauma, viral, bacterial, fungal, protozoal, and parasitic
CC infections. They can also treat disorders such as e.g., Alzheimer's
CC disease or a stroke. The NOVX encoding nucleic acids are useful for
CC expressing the NOVX proteins, to detect NOVX mRNA, or a genetic lesion in
CC a NOVX gene and to modulate NOVX activity. NOVX sequences are also useful
CC for identifying a cell or tissue type in a biological sample, to amplify
CC DNA sequences from very small biological samples such as tissues e.g.
CC hair or skin or body fluids in forensic biology and as primers and probes
CC for use in identifying and/or cloning NOVX homologues in other cell
CC types. The NOVX proteins are useful as an immunogen to generate
CC antibodies which are useful for diagnostically monitoring protein levels
CC and modulating NOVX activity. Cells comprising NOVX nucleic acids are
CC useful for producing non-human transgenic animals which are useful for
CC studying the function and/or activity of NOVX protein and for identifying
CC and/or evaluating modulators of NOVX protein activity. The NOVX nucleic
CC acids can be used in gene therapy. This sequence represents a NOVX
CC protein of the invention

XX

SQ Sequence 1450 AA;

Query Match 36.0%; Score 1297.5; DB 6; Length 1450;
Best Local Similarity 43.2%; Pred. No. 3.3e-52;
Matches 221; Conservative 47; Mismatches 216; Indels 27; Gaps 6;

Qy	94	CCPGFYESGEMCVPHCADKCVHG-----RCIAPNT-----CQCEPGWG GTNCS	136
Db	800	CLPGFVGS-----RCQDVCPAGWYGPSCQTRCSCANDGHCHPATGHCS CAPGWTGFSCQ	853
Qy	137	SACDGDHWGPHCTSRCQCKNG-ALCNPITGACHCAAGFRGWRCEDRCEQGT YGNDCHQRC	195
Db	854	RACDTGHWGPDCSHPCNCSAGHGSCDAISGLCLCEAGYVGPRCEQQCPQGHFGPGCEQLC	913

Qy 196 QCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPQCEQRCPCQNGGVCHHVTGECSC 255
 |||:| | | | :| | | | :| | | | | | | | | | | | | | | |
 Db 914 QCQHGAACDHVSGACTCPAGWRGTFCEHACPAGFFGLDCRSACNCTAGAACDAVNGSCLC 973
 Qy 256 PSGWMGTVCGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSPGYTGERCQDECPVGT 315
 |:| | | :| | | :| | | | | | | | | | :| | | :| | | | |
 Db 974 PAGRRGPRCAETCPAGLYGDNCRHSCLCQNGGTCDPVSGHACACPEGWAGLACEKECPPRD 1033
 Qy 316 YGVLCAETCQCVNNGGKCYHVSGACLCEAGFAGERCEARLCPEGLYGIKCDKRCPCHELENT 375
 | : | :| | | :| | | | | :| :| :| :| :| | | :| | |
 Db 1034 VRAGCRHSGGCLNGGLCDPHTGRCLCPAGWAGDKCQSP-CLRGWPGEACAQHCSC--PPG 1090
 Qy 376 HSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCDSVTGKCTCAPGFK 435
 :| | :| | | | :| | :| | | | :| :| | | | | :| | | | | |
 Db 1091 AACHHVTGACRCPPGFTGSGCEQGCPPGRYGPGEQQLCGCLNGGSCDAATGACRCPTGFL 1150
 Qy 436 GIDCSTPCPLGTYGINCSSRCGCKNDAVCSFVDGSCCTCKAGWHGVDCSIRCPSGTWGFGC 495
 | | | :| | | :| | | | | | | | | | | | | | | :| | |
 Db 1151 GTDCNLTCPQGRFGPNCTHVCGCGQGAACDPVTGTCLCPPGRAGVRCERGCPQNRFGVGC 1210
 Qy 496 NLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPCQDGTYGLNCAERCDCSHADGCHPTTG 555
 | | ! | | | | :| :| :| | | | | | | | | | | | :| | | | |
 Db 1211 EHTCSCRNGGLCHASNGSCSCGLGWTGRHCELACPPGRYGAACHLECSCHNNSTGEPATG 1270
 Qy 556 HCRCLPGWSGVHCDSVCAEGRWGPNCSLPCY 586
 | | | | :| | :| | | | | | :|
 Db 1271 TCRCPGFYQACEHPCPPGFHAGCQGLCW 1301

Search completed: March 26, 2004, 16:08:54
 Job time : 38.6674 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2004, 16:06:56 ; Search time 13.241 Seconds
(without alignments)
2284.780 Million cell updates/sec

Title: US-10-092-390-4
Perfect score: 3601
Sequence: 1 MVISLNSCLSFICLLCHWI.....HCDSVCAEGRWGPNCSLPCY 586

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	779	21.6	299	3	US-09-188-930-332	Sequence 332, App
2	779	21.6	299	4	US-09-312-283C-192	Sequence 192, App
3	779	21.6	299	4	US-09-312-283C-332	Sequence 332, App
4	759	21.1	299	3	US-09-188-930-192	Sequence 192, App
5	719	20.0	2523	1	US-08-185-432-18	Sequence 18, Appl
6	719	20.0	2523	4	US-08-899-232-3	Sequence 3, Appli
7	697	19.4	2321	4	US-09-230-652-2	Sequence 2, Appli
8	678	18.8	1193	2	US-08-400-159-10	Sequence 10, Appl
9	678	18.8	1193	3	US-08-611-729A-10	Sequence 10, Appl
10	676	18.8	2703	1	US-08-185-432-19	Sequence 19, Appl
11	676	18.8	2703	4	US-08-899-232-4	Sequence 4, Appli

12	666.5	18.5	2471	1	US-08-185-432-16	Sequence 16, Appl
13	666.5	18.5	2471	1	US-08-083-590A-19	Sequence 19, Appl
14	666.5	18.5	2471	3	US-08-532-384-19	Sequence 19, Appl
15	666.5	18.5	2471	4	US-08-899-232-1	Sequence 1, Appli
16	663	18.4	2556	1	US-08-185-432-17	Sequence 17, Appl
17	663	18.4	2556	4	US-08-899-232-2	Sequence 2, Appli
18	662	18.4	2556	1	US-08-083-590A-20	Sequence 20, Appl
19	662	18.4	2556	3	US-08-532-384-20	Sequence 20, Appl
20	659	18.3	1964	4	US-09-467-997-1	Sequence 1, Appli
21	644.5	17.9	2200	4	US-09-796-575-2	Sequence 2, Appli
22	636	17.7	1810	4	US-08-793-273C-4	Sequence 4, Appli
23	636	17.7	1810	5	PCT-US95-11684-4	Sequence 4, Appli
24	635	17.6	2199	4	US-08-793-273C-2	Sequence 2, Appli
25	635	17.6	2199	5	PCT-US95-11684-2	Sequence 2, Appli
26	634.5	17.6	1148	3	US-08-882-046-4	Sequence 4, Appli
27	634	17.6	1055	3	US-09-214-278-2	Sequence 2, Appli
28	634	17.6	1055	4	US-09-855-722-2	Sequence 2, Appli
29	633	17.6	1065	2	US-08-400-159-8	Sequence 8, Appli
30	633	17.6	1212	3	US-09-214-278-3	Sequence 3, Appli
31	633	17.6	1212	4	US-09-855-722-3	Sequence 3, Appli
32	633	17.6	1257	3	US-08-611-729A-8	Sequence 8, Appli
33	629	17.5	1010	3	US-08-882-046-7	Sequence 7, Appli
34	629	17.5	1036	4	US-09-068-740A-6	Sequence 6, Appli
35	629	17.5	1187	4	US-09-068-740A-7	Sequence 7, Appli
36	629	17.5	1208	4	US-09-199-865-1	Sequence 1, Appli
37	629	17.5	1218	2	US-08-400-159-6	Sequence 6, Appli
38	629	17.5	1218	3	US-08-611-729A-6	Sequence 6, Appli
39	629	17.5	1218	3	US-08-882-046-2	Sequence 2, Appli
40	629	17.5	1218	4	US-09-068-740A-11	Sequence 11, Appl
41	622	17.3	1238	3	US-09-214-278-5	Sequence 5, Appli
42	622	17.3	1238	4	US-09-855-722-5	Sequence 5, Appli
43	619.5	17.2	1248	3	US-08-882-046-6	Sequence 6, Appli
44	619	17.2	1218	3	US-09-214-278-7	Sequence 7, Appli
45	619	17.2	1218	4	US-09-855-722-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1

US-09-188-930-332

; Sequence 332, Application US/09188930A

; Patent No. 6150502

; GENERAL INFORMATION:

; APPLICANT: Watson, James D.

; APPLICANT: Strachan, Lorna

; APPLICANT: Sleeman, Matthew

; APPLICANT: Onrust, Rene

; APPLICANT: Murison, James Greg

; TITLE OF INVENTION: Compositions Isolated From Skin Cells

; TITLE OF INVENTION: and Methods For Their Use

; FILE REFERENCE: 11000.1011c1

; CURRENT APPLICATION NUMBER: US/09/188,930A

; CURRENT FILING DATE: 1998-11-09

; NUMBER OF SEQ ID NOS: 348

; SOFTWARE: FastSEQ for Windows Version 3.0

; SEQ ID NO 332

; LENGTH: 299
; TYPE: PRT
; ORGANISM: Mouse
US-09-188-930-332

Query Match 21.6%; Score 779; DB 3; Length 299;
Best Local Similarity 39.9%; Pred. No. 7e-43;
Matches 127; Conservative 27; Mismatches 118; Indels 46; Gaps 1;

```
Qy      165 GACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDL 224
          |||:| ||| | | | | | :| | | | | || | :| | |||| || ||
Db      4   GACYCPAGFLGADCSLACPGQGRFGPSCAHVCTCGQGAACDPVSGTCICPPGKTGGHCERG 63

Qy      225 CPPGKHGPQCEQRCPCQNGGVCHHVTGECSCPSGWMGTVCQGQPCPEGRFGKNCSEQECQCH 284
          || :| | | :| | :|||:| | | | | ||| | | | | | | :| | || |
Db      64 CPQDRFGKGCEHKCACRNGGLCHATNGSCSCPLGWMGPHCEHACPAGRYGAACLLECSQ 123

Qy      285 NGGTCDAAATGQCHCSPGYTGERCQDECPVGTYGVLCAETCQCVNGGKCYHVSAGACLCEAG 344
          | |::: :| | | ||: | :| | | :| | | :| | | | | || || ||
Db      124 NNGSCEPTSGACLCPGPGFYGQACEDTCPAGFHGSGCQRVCECQQGAPCDPVSGRCLCPAG 183

Qy      345 FAGERCEARLCPEGLYGIKCDKRCPCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGF 404
          | |: :| | | | | | | | | | | | | | | | | | | | | | | |
Db      184 FRGQ-----FCERGCKPGF 197

Qy      405 YGEACQQICSCQNGADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVC 464
          :|: | | |:| | | | :| | | | | | | | | | :| | | | | |
Db      198 FGDGCLQQCNCPTGVPCDPISGLCLCPPGRAGTTCDLDCRRGRFGPGCALRCDCGGGADC 257

Qy      465 SPVDGSCTCKAGWHGVDC 482
          |: | | | :| | |
Db      258 DPISGQCHCVDSYTGPTC 275
```

RESULT 2

US-09-312-283C-192
; Sequence 192, Application US/09312283C
; Patent No. 6573095
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated from Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c2
; CURRENT APPLICATION NUMBER: US/09/312,283C
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 425
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 192
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Mouse
US-09-312-283C-192

Query Match 21.6%; Score 779; DB 4; Length 299;
 Best Local Similarity 39.9%; Pred. No. 7e-43;
 Matches 127; Conservative 27; Mismatches 118; Indels 46; Gaps 1;

```

Qy      165 GACHCAAGFRGWRCEDRCEQGTGYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDL 224
      |||:| ||| | | | || :| | | | | | | | | | | | | |
Db      4 GACYCPAGFLGADCSLACPQGRFGPSCAHVCTCGQGAACDPVSGTCICPPGKTGGHCERG 63

Qy      225 CPPGKHGPQCEQRCPCQNGGVCHHVTGECSCPSGWMGTVCGQPCPEGRFGKNCSQECQCH 284
      || : | || :| | :|||:| | ||| ||| | | |||:| | || |
Db      64 CPQDRFGKGCEHKACRNGGLCHATNGSCSCPLGWMGPHCEHACPAGRYGAACLLECSQ 123

Qy      285 NGGTCDAAATGQCHCSPGYTGERCQDECPVGTYGVLCAETCQCVNGGKCYHVSAGACLCEAG 344
      | |:| :| | | ||: | :| | | :| | | :| | | ||| ||| ||
Db      124 NNGSCEPTSGACLCGPGFYGQACEDTCPAGFHGSGCQRVCECQQGAPCDFVSGRCLCPAG 183

Qy      345 FAGERCEARLCPEGLYGIKCDKRCPCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGF 404
      | |: | | :| | | | | | | | | | | | | | | | | |
Db      184 FRGQ-----FCERGCKPGF 197

Qy      405 YGEACQQICSCQNGADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVC 464
      |:| | | |:| | | || :| | | | | | | | | | | | | |
Db      198 FGDGCLQQCNCPTGVPCDPISGLCLCPPGRAGTTCDLDCRRGRFGPGCALRCDCGGGADC 257

Qy      465 SPVDGSCTCKAGWHGVDC 482
      |:| | | :| | |
Db      258 DPISGQCHCVDSYTGPTC 275

```

RESULT 3

US-09-312-283C-332
 ; Sequence 332, Application US/09312283C
 ; Patent No. 6573095
 ; GENERAL INFORMATION:
 ; APPLICANT: Watson, James D.
 ; APPLICANT: Strachan, Lorna
 ; APPLICANT: Sleeman, Matthew
 ; APPLICANT: Onrust, Rene
 ; APPLICANT: Murison, James G.
 ; APPLICANT: Kumble, Krishanand D.
 ; TITLE OF INVENTION: Compositions Isolated from Skin Cells
 ; TITLE OF INVENTION: and Methods for Their Use
 ; FILE REFERENCE: 11000.1011c2
 ; CURRENT APPLICATION NUMBER: US/09/312,283C
 ; CURRENT FILING DATE: 1999-05-14
 ; NUMBER OF SEQ ID NOS: 425
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 332
 ; LENGTH: 299
 ; TYPE: PRT
 ; ORGANISM: Mouse
 US-09-312-283C-332

Query Match 21.6%; Score 779; DB 4; Length 299;
 Best Local Similarity 39.9%; Pred. No. 7e-43;
 Matches 127; Conservative 27; Mismatches 118; Indels 46; Gaps 1;

Qy 165 GACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDL 224
 |||:| ||| | | | | :| | | | | | | :| | ||| | | ||
 Db 4 GACYCPAGFLGADCSLACPQGRFGPSCAHVCTCGQGAACDPVSGTCICPPGKTGGHCERG 63

Qy 225 CPPGKHGPPQCEQRCPCQNGGVCHHVTGECSCPSGWMGTVCQGQPCPEGRFGKNCSQECQCH 284
 || :| | | :| | :|||:| | | | | | | | | | | | :| | | | |
 Db 64 CPQDRFGKGCEHKCACRNGGLCHATNGSCSCPLGWMGPHCEHACPAGRYGAACLLECSQ 123

Qy 285 NGGTCDAAATGQCHCSPGYTGERCQDECPVGTYGVLCAETCQCVNGGKCYHVSGACLCEAG 344
 | |:| :| | | | |:| :| | | :| | | :| | | | | | | |
 Db 124 NNGSCEPTSGACLCGPGFYGQACEDTCPAGFHGSGCQRVCECQQGAPCDPVSGRCLCPAG 183

Qy 345 FAGERCEARLCPEGLYGIKCDKRCPCCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGF 404
 | |:| :| | | | | :| | | | | | | | | | | | | | | |
 Db 184 FRGQ-----FCERGCKPGF 197

Qy 405 YGEACQQICSCQNGADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVC 464
 |:| | | |:| | | | |:| | | | | | | | | | | | | | | |
 Db 198 FGDGCLQQCNCPTGVPCDPISGLCLCPPGRAGTTCDLDCRRGRFGPGCALRCDCGGGADC 257

Qy 465 SPVDGSCCTCKAGWHGVDC 482
 |:| | | | :| | |
 Db 258 DPISGQCHCVDSYTGPTC 275

RESULT 4

US-09-188-930-192

; Sequence 192, Application US/09188930A

; Patent No. 6150502

; GENERAL INFORMATION:

; APPLICANT: Watson, James D.

; APPLICANT: Strachan, Lorna

; APPLICANT: Sleeman, Matthew

; APPLICANT: Onrust, Rene

; APPLICANT: Murison, James Greg

; TITLE OF INVENTION: Compositions Isolated From Skin Cells

; TITLE OF INVENTION: and Methods For Their Use

; FILE REFERENCE: 11000.1011c1

; CURRENT APPLICATION NUMBER: US/09/188,930A

; CURRENT FILING DATE: 1998-11-09

; NUMBER OF SEQ ID NOS: 348

; SOFTWARE: FastSEQ for Windows Version 3.0

; SEQ ID NO 192

; LENGTH: 299

; TYPE: PRT

; ORGANISM: mouse

; FEATURE:

; NAME/KEY: UNSURE

; LOCATION: (98)...(98)

; NAME/KEY: UNSURE

; LOCATION: (239)...(239)

US-09-188-930-192

Query Match 21.1%; Score 759; DB 3; Length 299;

Best Local Similarity 39.3%; Pred. No. 1.3e-41;

Matches 125; Conservative 27; Mismatches 120; Indels 46; Gaps 1;

Qy 165 GACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDL 224
 |||:| ||| | | || :| | | || || :| | ||| || ||
 Db 4 GACYCPAGFLGADCSLACPQGRFGPSCAHVCTCGQGAACDPVSGTCICPPGKTGGHCERG 63

Qy 225 CPPGKHGPPQCEQRCPCQNGGVCHHVTGECSCPSGWMGTVCQGQPCPEGRFGKNCSQECQCH 284
 || :| || :| ||:|||| | |||| | | | |||:| | || |
 Db 64 CPQDRFGKGCEHKCACRNGGLCHATNGSCSCPLGXMGPHCEHACPAGRYGAACLLECSCQ 123

Qy 285 NGGTCDAATGQCHCSPGYTGERCQDECPVGTYGVLCAETCQCVNNGGKCYHVSGACLCEAG 344
 | |:| :| | | |:| |:| || | :| | |:| | | ||| ||| ||
 Db 124 NNGSCEPTSGACLCPGPGFYGQACEDTCPAGFHGSGCQRVCECQQGAPCDPVSGRCLCPAG 183

Qy 345 FAGERCEARLCPEGLYGIKCDKRCPCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGF 404
 | |:| :| | | ||| :| | |||
 Db 184 FRGQ-----FCERGCKPGF 197

Qy 405 YGEACQQICSCQNGADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVC 464
 |:| | | |:| | || :| | | | | | | | :| | |:| | | | |
 Db 198 FGDGCLQQCNCPTGVPCDPISGLCLCPPGRAGTTCDLDCRRXRFGPGCALRCDGCGGADC 257

Qy 465 SPVDGSCTCKAGWHGVDC 482
 |:| | | :| | |
 Db 258 DPISGQCHCVDSYTGPTC 275

RESULT 5

US-08-185-432-18

; Sequence 18, Application US/08185432

; Patent No. 5750652

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, Spyridon

; APPLICANT: Busseau, Isabelle

; APPLICANT: Diederich, Robert J.

; APPLICANT: Xu, Tian

; APPLICANT: Matsuno, Kenji

; TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND

; TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PENNIE & EDMONDS

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036-2711

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/185,432

; FILING DATE: 21-JAN-1994

; CLASSIFICATION: 530

; ATTORNEY/AGENT INFORMATION:

; NAME: Misrock, S. Leslie

```

;      REGISTRATION NUMBER: 18,872
;      REFERENCE/DOCKET NUMBER: 7326-006
;      TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (212) 790-9090
;      TELEFAX: (212) 869-8864/9741
;      TELEX: 66141 PENNIE
;      INFORMATION FOR SEQ ID NO: 18:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH: 2523 amino acids
;      TYPE: amino acid
;      TOPOLOGY: unknown
;      MOLECULE TYPE: protein
US-08-185-432-18

```

```

Query Match          20.0%; Score 719; DB 1; Length 2523;
Best Local Similarity 25.9%; Pred. No. 3.2e-38;
Matches 225; Conservative 60; Mismatches 221; Indels 364; Gaps 51;

```

```

Qy      5 LNSCLSFICL-----LLCHWIGTASPLNLED-----PNVCSHW-----ESYS 41
      :| ||| ||           :|       :| |       |:| :       :|
Db      603 INECLSKPCLNGGQCTDRENGYICTCPKGTGVCNCTKIDDCASNLCNDNGKCIDKIDGYE 662

Qy      42 VTVQESYPHPFDQIYYT-----SCTDILNWFKCTRHRVSYRTAYRHGEKTMYYR 90
      | : | | |           :| | :| | |
Db      663 CTCEPGYTGKLCNININECDNPNCRNGGTCKDQINGFTCV----- 702

Qy      91 KSQCCPGFYESGEMC---VPHC-ADKCVHGRC---IAPNTCQCEPGWGGTNC---SSACD 140
      || | | | | | :| :| | | : | | | | :| | :| :|
Db      703 ----CPDGYHD-HMCLSEVNECNPNPCIHGACHDGVNGYKCDCEAGWSGSNCDINNNECE 757

Qy      141 GDHWGPHCTSRCQCKNGALCNPITGA--CHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCQ 198
      : | | | | :| | | | | | | : | | :| | |
Db      758 SN-----PCMNGGTCKDMTGAYICTCKAGFSGPNCQ-----TNINECSSN-PCL 800

Qy      199 NGATC-DHVTG-ECRCPPGYTGAFCEDLCPGKHGPQCEQRC---PCQNGGVCHH----V 249
      | | | | | :| | | | | : | | | | | :| | | |
Db      801 NHGTCIDDVAGYKCNCLPYTGAICEAVLAP-----CAGSPCKNGGRCKESEDFFE 850

Qy      250 TGECSGPSGWMGTVCQGQPCPEGRFGKNCSQEC---QCHNGGTCDAAATG--QCHCSPGYTG 304
      | | | | | | | : | | | | | | | | :| :| | | | |
Db      851 TFSCECPPGWQGGQTC-----EIDMNECVNRPCRNATCQNTNGSYKCNCKPGYTG 900

Qy      305 ERCQ---DECPVGTYGVLCAETCQCVNGGKCYHVSQA--CLCEAGFAGERCEARL----- 354
      |: | :| : | | | | | | | | | | :| :| :| :|
Db      901 RNCEMDIDDC-----QPNPCHNGGSCSDGINMFFCNCPCAGFRGPKCEEDINECAS 950

Qy      355 -----CPEGLYGIKCDKRCPCHE----NTHSCHPMSG----ECAC 387
      | | | | : | | | | :| :| | | |
Db      951 NPCKNGANCTDCVNSYTCTCQPGFSGIHCESNTPDCTESSCFNGGTC--IDGINTFTCQC 1008

Qy      388 KPGWSGLYC---NE-----TCSPGFYGEACQOI---CS---CQ 416
      ||:| | | | | | | | | :| | | : | | :|
Db      1009 PPGFTGSYCQHDINECDSPCLNGGTCQDSYGTYKCTCPQGYTGLNQNLRWCDSSPCK 1068

Qy      417 NGADCDSVTG--KCTCAPGFKGIDCSTP----- 442
      || | :| | :| :| | |
Db      1069 NGGKCWQTNNFYRCECKSGWTGVYCDVPSVSCEVAAKQQGVDIVHLCRNSGMCVDTGNTH 1128

```


Qy 443 ---CPLGTYGINCSSR---CG---CKNDAVCSPVDG--SCTCKAGWHGVDCS----- 483
 | | | | : | : | | : | | | | : | | : |
 Db 1129 FCRCQAGYTGSYCEEQVDECSNPNCQNGATCTDYLGYSCECVAGYHGVCNSEEINECLS 1188
 Qy 484 -----IRCPSTWGFNCNLT---C-----QCLNGGACNTLDG 512
 | | | | : | : | | | : | | | | |
 Db 1189 HPCQNGGTCIDLINTYKCSNPRGTQGVHCEINVDDCTPFYDSFTLEPKCFNNGKCIDRVG 1248
 Qy 513 --TCTCAPGWRGEKCE-----LPCQD-GTYGLNCAE-----RCDC-----SH 546
 | | | : | : | | | | : | : | |
 Db 1249 GYNICPPGFGVGERCEGDVNECLSNPCDSRGTQ--NCIQLVNDYRCECRQGFTGRRCESV 1306
 Qy 547 ADGC-----HPTTGH-CRCLPGWSGVHCD----- 569
 | | | : | | : | | : | | :
 Db 1307 VDGCKGMPCRNGGTCAVASNTERGFICKPPGFDGATCEYDSRTCSNLRQNGGTCISVL 1366
 Qy 570 ----SVCAEGRWGPNC-----SLPCY 586
 | : | | | | | | | | | |
 Db 1367 TSSKCVCSGYTGATCQYPVISPCASHPCY 1396

RESULT 6

US-08-899-232-3

; Sequence 3, Application US/08899232
 ; Patent No. 6436650
 ; GENERAL INFORMATION:
 ; APPLICANT: Artavanis-Tsakonas, Spyridon
 ; APPLICANT: Qi, Huilin
 ; TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON
 ; FILE REFERENCE: 7326-046
 ; CURRENT APPLICATION NUMBER: US/08/899,232
 ; CURRENT FILING DATE: 1997-07-23
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 3
 ; LENGTH: 2523
 ; TYPE: PRT
 ; ORGANISM: Xenopus sp.
 US-08-899-232-3

Query Match 20.0%; Score 719; DB 4; Length 2523;
 Best Local Similarity 25.9%; Pred. No. 3.2e-38;
 Matches 225; Conservative 60; Mismatches 221; Indels 364; Gaps 51;

Qy 5 LNSCLSFICL-----LLCHWIGTASPLNLED-----PNVCSHW-----ESYS 41
 : | | | | : | : | | : | : : |
 Db 603 INECLSKPCLNGGQCTDRENGYICTCPKGTGVCNCTKIDDCASNLCNDNGKCIDKIDGYE 662
 Qy 42 VTVQESYPHPFDQIYYT-----SCTDILNWFKCTRHRVSYRTAYRHGEKTMYYR 90
 | : | | : | | : | | | |
 Db 663 CTCEPGYTGKLCNININECDSNPCRNGGTCCKDQINGFTCV----- 702
 Qy 91 KSQCCPGFYESGEMC---VPHC-ADKCVHGRC---IAPNTCQCEPGWGGTNC---SSACD 140
 | | | | | : | : | | : | : | : | : | :
 Db 703 ----CPDGYHD-HMCLSEVNECNPNCTIHGACHDGVNGYKCDCEAGWSGSNCDINNNECE 757

QY 141 GDHWGPHCTSRCQCKNGALCNPITGA--CHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCQ 198
 : | | | : | | | | : | : |
 Db 758 SN-----PCMNGGTCKDMTGAYICTCKAGFSGPNCQ-----TNINECSSN-PCL 800
 QY 199 NGATC-DHVTG-ECRCPPGYTGAFCEDLCPGKHGPQCEQRC---PCQNGGVCHH----V 249
 | | | | : | | | | : | | : | |
 Db 801 NHGTCIDDVAGYKNCNMLPYTGAICEAVLAP-----CAGSPCKNGGRCKESEDFE 850
 QY 250 TGECSGPSGWMGTVCQGQPCPEGRFGKNCSQEC---QCHNGGTCDAAATG--QCHCSPGYTG 304
 | | | | | : | | | | | : | | | |
 Db 851 TFSCECPPGWQGGQTC-----EIDMNECVNRPCRNGATCQNTNGSYKCNCKPGYTG 900
 QY 305 ERCQ---DECPVGTYGVLCAETCQCVNNGGKCYHVSGA--CLCEAGFAGERCEARL----- 354
 | : | : : | | | | | | | : | :
 Db 901 RNCMDIDDC-----QPNPCHNGGSCSDGINMFFCNCAPAGFRGPKCEEDINECAS 950
 QY 355 -----CPEGLYGIKCDKRCPCBLE----NTHSCHPMSG----ECAC 387
 | | | | : | | | : | : |
 Db 951 NPCKNGANCTDCVNSYTCTCQPGFSGIHCESNTPDCTESSCFNGGTC--IDGINTFTTCQC 1008
 QY 388 KPGWSGLYC---NE-----TCSPGFYGEACQOI---CS---CQ 416
 | : : | | | | | : | | : | :
 Db 1009 PPGFTGSYCQHDINECDSPCLNGGTCQDSYGYTKCTCPQGYTGLNCQNLVRWCDSSPCK 1068
 QY 417 NGADCDSVTG--KCTCAPGFGKIDCSTP----- 442
 | | : | | : | : |
 Db 1069 NGGKCWQTNNFYRCECKSGWTGVYCDVPSVSCEVAAKQQGVDIVHLCRNSGMCVDTGNTH 1128
 QY 443 ---CPLGTYGINCSSR---CG---CKNDVAVCSVPDG--SCTCKAGWHGVDCS----- 483
 | | | : | : | : | | | : | : | : |
 Db 1129 FCRCQAGYTGSYCEEQVDECSNPNCQNGATCTDYLGGYSCECVAGYHGVNCSEEINECLS 1188
 QY 484 -----IRCPSTWGFGCNLT---C-----QCLNGGACNTLDG 512
 | | | | : | : | | |
 Db 1189 HPCQNGGTCLIDLINTYKCS CPRGTQGVHCEINVDDCTPFYDSFTLEPKCFNNGKCIDRVG 1248
 QY 513 --TCTCAPGWRGEKCE-----LPCQD-GTYGLNCAE-----RCDC-----SH 546
 | | | : | : | | | | : | : |
 Db 1249 GYNICPPPGFVGERCEGDVNECLSNPCDSRGTO--NCIQLVNDYRCECRQGFTGRRCESV 1306
 QY 547 ADGC-----HPTTGH-CRCLPGWSGVHCD----- 569
 | | : | | : | : | :
 Db 1307 VDGCKGMPCRNGGTCAVASNTERGFICKCPPGFDGATCEYDSRTCSNLRQNGGTCLISVL 1366
 QY 570 ----SVCAEGRWGPNC-----SLPCY 586
 | : | | | | |
 Db 1367 TSSKVCVSEGYTGATCQYPVISPCASHPCY 1396

RESULT 7

US-09-230-652-2

; Sequence 2, Application US/09230652A

; Patent No. 6537775

; GENERAL INFORMATION:

; APPLICANT: Tournier-Lasserre, Elisabeth

; APPLICANT: Joutel, Anne

; APPLICANT: Bousser, Marie-Germaine

```

; APPLICANT: Bach, Jean-Francois
; TITLE OF INVENTION: GENE INVOLVED IN CADASIL, METHOD OF DIAGNOSIS AND
; TITLE OF INVENTION: THERAPEUTIC APPLICATION
; FILE REFERENCE: 03715.0048-00000
; CURRENT APPLICATION NUMBER: US/09/230,652A
; CURRENT FILING DATE: 1999-05-17
; EARLIER APPLICATION NUMBER: FR 96 09733
; EARLIER FILING DATE: 1996-08-01
; EARLIER APPLICATION NUMBER: FR 97 04680
; EARLIER FILING DATE: 1997-04-16
; EARLIER APPLICATION NUMBER: PCT/FR97/01433
; EARLIER FILING DATE: 1997-07-31
; NUMBER OF SEQ ID NOS: 163
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2321
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human ADNc No. 6537775ch 3
US-09-230-652-2

```

```

Query Match          19.4%; Score 697; DB 4; Length 2321;
Best Local Similarity 25.2%; Pred. No. 7.6e-37;
Matches 226; Conservative 61; Mismatches 249; Indels 360; Gaps 51;

```

```

Qy      5 LNSCLS-----FICLLLCHWIGTASPLNLED-----PNVC-SHWESY 40
      :| |||          | |: : : || ::::          || :
Db      432 VNECLSGPCRNOATCLDRIGQFTCICMAGFTGTYCEVDIDECQSSPCVNGGVCKDRVNGF 491

Qy      41 SVTVQESYPHPFDQIYYTSC--TDILNWFKCTRHRVSYRTAYRHGEKTMYYRKSQCCPGF 98
      | | : : | | | | | | | | : : | |
Db      492 SCTCPSGFGSGSTCQLDVDECASTPCRNGAKCVDQPDGY-----ECRCAEGF 537

Qy      99 YESGEMC---VPHCA-DKCVHGRC---IAPNTCQCEPGWGGTNCSSACDGDHWGPHCTSR 151
      | :| | | : | | ||| | | : | | | : | | | | | | | | :
Db      538 --EGTLCDRNVDDCSPDPCHHGRCVDGIASFSCACAPGYTGTRCESQVD-----ECRSQ 589

Qy      152 CQCKNGALCNPITG--ACHCAAGFRGWRCE---DRCEQG--TYG--NDCHQR--CQCQNG 200
      |::| | : | | : | | | | | | | | : | | | | | | |
Db      590 -PCRHHGKCLDLVDKYLRCRPSGTTGVNCEVNIDDCASNPCFTFGVCRDGINRYDCVCQPG 648

Qy      201 AT---CDHVTGECRCPPGYTGAFCD-----LCPPGKHGPQC---EQRC---PCQNGG 244
      | | : | | | | | | | | | | | | | | | | | : |
Db      649 FTGPLCNVEINECASSPCGEGGSCVDGENGFRLCPPGSLPPLCLPPSHPCAHEPCSH-G 707

Qy      245 VCHHVTG--ECSCPSGWMGTVCQG-----PCPEGRFGK 275
      :|: | | | | | | | | | | | | | | :
Db      708 ICYDAPGGFRCVCEPGWSGPRCSQSLARDACESQPCRAGGTCSSDGMGFHCTCPPGVQGR 767

Qy      276 NCS--QEC---QCHNGGTCDAAATGQ---CHCSPGYTGERCQ---DEC----PVGTYGVLC 320
      | | | : || |::| | | | | : | | | | | | | | : |
Db      768 QCELLSPCTPNPCEHGGRCESAPQLPVCSCFPQGWQGPRCQQDVDECAGPAPCGPHGI-C 826

Qy      321 AE-----TCQ-----CVNGGKCYHVSG--ACLCEAGFAGERCEA 352
      ||          |::| | | | : | | | | | | | |
Db      827 TNLAGSFSCCTCHGGYTGPSCDQDINDCDPNPCLNGGSCQDGVGSFSCSCLPGFAGPRC-A 885

```

Qy	353	R-----LCPEGLYGIKCDKRCPCHELENTSCHPM-----	381
Db	886	RDVDECLSNPCGPGTCTDHVASFTCTCPPGYGGFHCEQDLP-----DCSPSSCFNGGT	938
Qy	382	-----SGECACKPGWSGLYCNE-----TCSPGFYGEACQ	410
Db	939	CVDGVNSFSCLCRPGYTGAHCQHEADPCLSRPCLHGGVCSAAHPGFRCTCLESTGPQCQ	998
Qy	411	QI---CS---CQNGADCDSVTGKCTCAPGFKGIDC---STP-----	442
		:	
Db	999	TLVDWCSRQPCQNGGRCVQTGAYCLCPPGWSGRLCDIRSLPCREAAQIGVRLEQLCQAG	1058
Qy	443	-----CPLGTYGINCSSRCG-----CKNDVAVCSFVDGS--CTCKAGWHGVD	481
Db	1059	GQCVDEDESSHVCVCEPGRGTGSHCEQEVDPCLAQPCQHGGTCRGYMGGYMCECLPGYNGDN	1118
Qy	482	CS-----IRCPSGTWGFGCNLT---C-----QCLN	503
Db	1119	CEDDVDECASQPCQHGGSCIDLVARVLCSCPPGTGLVLCINEDDCGPGPPLDSGPRCLH	1178
Qy	504	GGACNTLDG--TCTCAPGWRGEKCEL---PCQDGTYGLNCAERCDCSHADGCHPTTG---	555
Db	1179	NGTCVDLVGGFRCTCPPGYTGLRCEADINECRSGA-----CHAAHTRDCLQDPGGGF	1230
Qy	556	HCRCLPGWSGVHCDSV-----CAEGRWGPNC	581
Db	1231	RCLCHAGFSGPRCQTVLSPCESQPCQHGGQCRPSPGPGGGLTFTCHCAQPFWGPRC	1286

RESULT 8

US-08-400-159-10

; Sequence 10, Application US/08400159

; Patent No. 5869282

; GENERAL INFORMATION:

; APPLICANT: Ish-Horowicz, David
 ; APPLICANT: Henrique, Domingos M.P.
 ; APPLICANT: Lewis, Julian H.
 ; APPLICANT: Myat, Anna M.
 ; APPLICANT: Fleming, Robert J.
 ; APPLICANT: Artavanis-Tsakonas, Spyridon
 ; APPLICANT: Mann, Robert S.
 ; APPLICANT: Gray, Grace E.

; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE

; TITLE OF INVENTION: SERRATE GENE AND METHODS BASED THEREON

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10036-2711

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS

```

;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/400,159
;   FILING DATE: 07-MAR-1995
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Misrock, S. Leslie
;   REGISTRATION NUMBER: 18,872
;   REFERENCE/DOCKET NUMBER: 7326-029
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 790-9090
;   TELEFAX: (212) 869-9741/8864
;   TELEX: 66141 PENNIE
;   INFORMATION FOR SEQ ID NO: 10:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1193 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-400-159-10

```

```

Query Match          18.8%; Score 678; DB 2; Length 1193;
Best Local Similarity 27.4%; Pred. No. 6.8e-36;
Matches 197; Conservative 63; Mismatches 214; Indels 244; Gaps 46;

```

```

Qy      50 HPFDQIYYTSCTDILNWFKCTRHRVSYRTAYRHGEKTM-----YRRKSQCCPGFYESG 102
      | || :| : :| : :|| | : | : ||
Db      184 HTCDQNGNKTCLGWTGPECNKAICRQGCSPKHGSCTVPGECCRCQYGWQGQYC----- 236

Qy      103 EMCVPHCADKCVHGRCIAPNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCN- 161
      :|: || ||| || | || ||| || | :| : ||| |:
Db      237 DKCIPH--PGCVHGTCIEPWQCLCETNWGG----QLCDKDL--NYCGTHPPCLNGGTCSN 288

Qy      162 --PITGACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCQNGATC-DHVTG-ECRCPPGYT 217
      | | | :| || | : || || :| : || || |||:
Db      289 TGPDKYQCSCPEGYSQONCE-IAEHACLSDPCH-----NGGSCLETSTGFECVCAPGWA 341

Qy      218 GAFCEDL---CPPGKHGPQCEQRCPCQNGGVCHHVTG--ECSCPSGWMGTVC----- 264
      | || | | || :|| | : :|| | | | |
Db      342 GPTCTDNIDDCSPN-----PCGHGGTCQDLVDGFKCICPPQWTGKTCQLDANECE 391

Qy      265 GQPCPE-----GRFGKNCS---QEC--QCHNGGTC-DAATG-QCHCSP 300
      |: || | | | | :| || |||:| | | :| |||
Db      392 GKPCVNANSCRNLIGSYDCITGWSGHNCDININDCRGQCQNGGSCRDVNGYRCICSP 451

Qy      301 GYTGERCQ---DECPVGTYGVLCAETCQCVNNGGKCY-HVSG-ACLCEAGFAGERCEARL- 354
      || |: |: :|| :|: ||| | :| ||| |||:| |: :
Db      452 GYAGDHCEKDINEC-----ASNPCMNGGHCQDEINGFQCLCPAGFSGNLCQLDID 501

Qy      355 -----CPEGLYGIKCD-----KRCPCHLENT----- 375
      ||| | | : || : ::
Db      502 YCEPNPCQNGAQCFNLAMDYFCNCPEDYEGKNCSHLKDHCRTPCEVIDSCTVAVASNST 561

Qy      376 -----HSCHPMMSG---ECACKPGWSGLYCNETCSPGFYGEACQQICSCQN 417
      | :| | | |:| ||| : | : |:|
Db      562 PEGVRYISSNVCGPHGKCKSQAGGKFTCECNKGFTGTYPHENIND-----CES-NPCKN 614

```

Qy 418 GADC-DSVTG-KCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDVCSFV--DGSCTC 473
 | | | | | | : | : | | | | : | | |
 Db 615 GGTCTIDGVNSYKICISDGEWETYCET-----NINDCSKNPCHNGGTCRDLVNDFFCEC 667
 Qy 474 KAGWHGVDCSIR-----CPSGTWGFGCNL---TC---QC 501
 | | | | | | | | : | | | : | | |
 Db 668 KNGWKGTCHSRDSQCDEATCNNGGTCYDEGDTFKCMCPAGWEGATCNIARNSSCLPNPC 727
 Qy 502 LNGGACNTLDG---TCTCAPGWRGEKCEL-----PC-QDGTYGLNCAE-----RCDC 544
 | | | | : | | | | | | | | : | | |
 Db 728 HNGGTC-VVSGDSFTCVCKEGWEGPTCTQNTNDCSPHPCYNSGT----CVDGDNWYRCEC 782
 Qy 545 S---HADGCHPTTGHCR-----CLPGWSGVHCDVCAEGRWGPNC----SLPCY 586
 : | | : | : | : | | | | : | | :
 Db 783 APGFAGPDCRININECQSSPCAFGATCVDEINGYRC--ICPPGRSGPGCQEVTRPCF 838

RESULT 9

US-08-611-729A-10

; Sequence 10, Application US/08611729A

; Patent No. 6004924

; GENERAL INFORMATION:

; APPLICANT: Ish-Horowicz, David
 ; APPLICANT: Henrique, Domingos M.P.
 ; APPLICANT: Lewis, Julian H.
 ; APPLICANT: Myat, Anna M.
 ; APPLICANT: Fleming, Robert J.
 ; APPLICANT: Artavanis-Tsakonas, Spyridon
 ; APPLICANT: Mann, Robert S.
 ; APPLICANT: Gray, Grace E.

; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE

; TITLE OF INVENTION: SERRATE GENE AND METHODS BASED THEREON

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10036-2711

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/611,729A

; FILING DATE: 06-MAR-1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Misrock, S. Leslie
 ; REGISTRATION NUMBER: 18,872
 ; REFERENCE/DOCKET NUMBER: 7326-037

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 790-9090
 ; TELEFAX: (212) 869-9741/8864
 ; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1193 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-611-729A-10

Query Match 18.8%; Score 678; DB 3; Length 1193;
Best Local Similarity 27.4%; Pred. No. 6.8e-36;
Matches 197; Conservative 63; Mismatches 214; Indels 244; Gaps 46;

```
Qy      50 HPFDQIYYTSCTDILNWFKCTRHRVSYRTAYRHGEKTM-----YRRKSQCCPGFYESG 102
      | || :| : :| : :|| | : | |
Db      184 HTCDQNGNKTCLEGWTGPECNKAICRQGCSPKHGSCVTPGECRCQYGWQGQYC----- 236

Qy      103 EMCVPHCADKCVHGRCIAPNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCN- 161
      :|:| | | | | | | | | | | | :| : | | | :
Db      237 DKCIPH--PGCVHGTCTIEPWQCLCETNWGG----QLCDKDL--NYCGTHPPCLNGGTCSN 288

Qy      162 --PITGACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCQNGATC-DHVTG-ECRCPPGYT 217
      | | | | :| | | | : || | :| : | | | | | :
Db      289 TGPDKYQCSCEGYSGQNC-IAEHACLSDPCH-----NGGSCLETSTGFECVCAPGWA 341

Qy      218 GAFCEDL---CPPGKHGPQCEQRCPCQNGGVCHHVTG--ECSCPSGWMGTVC----- 264
      | | | | | | | | :|| | : :| | | | | |
Db      342 GPTCTDNIDDCSPN-----PCGHGGTCQDLVDGFKCICPPQWTGKTCQLDANECE 391

Qy      265 GQPCPE-----GREGKNCS---QEC--QCHNGGTC-DAATG-QCHCSP 300
      |:| | | | | | :| | | | | :| | | |
Db      392 GKPCVNANSCRNLIGSYDCITGWSGHNCNINDCRGQCQNGGSCRDLVNGYRCICSP 451

Qy      301 GYTGERCQ---DECPVGTYGVLCAETCQCVNCGKCY-HVSG-ACLCEAGFAGERCEARL- 354
      || |:| : :|| : |:|| | :| | | | | | : :
Db      452 GYAGDHCEKDINEC-----ASNPCMGGHCQDEINGFQCLCPAGFSGNLCQLDID 501

Qy      355 -----CPEGLYGIKCD-----KRCPCHLENT----- 375
      || | | | : || : :
Db      502 YCEPNPCQNGAQCFNLAMDYFCNCPEDYEGKNCSHLKDHCRTPCEVIDSCTVAVASNST 561

Qy      376 -----HSCHPMMSG---ECACKPGWSGLYCNETCSPGFYGEACQQICSCQN 417
      | :| | | | :| | | : | :|
Db      562 PEGVRYISSNVCGPHGKCKSQAGGKFTCECNKGFTGTYPHENIND-----CES-NPCKN 614

Qy      418 GADC-DSVTG-KCTCAPGFGKIDCSTPCPLGTYGINCSSRCGCKNDAVCSPV--DGSCTC 473
      | | | | | | :| :| | | | | : | | |
Db      615 GGTCIDGVNSYKICSDGWEGTYCET-----NINDCSKNPCHNGGTCRDLVNDFCEC 667

Qy      474 KAGWHGVDCSIR-----CPSGTWGFGCNL----TC---QC 501
      | | | | | | | | | | | | :| | |
Db      668 KNGWKGTCHSRDSQCDEATCNNGGTCYDEGDTFKCMCPAGWEGATCNIAARNSSCLPNPC 727

Qy      502 LGGACNTLDG---TCTCAPGWRGEKCEL-----PC-QDGTYGLNCAE-----RCDC 544
      || | | :| | | | | | | | | | : | | |
Db      728 HNGGTC-VVSGDSFTCVCKEGWEGPTCTQNTNDCSPHPCYNSGT---CVDGDNWYRCEC 782

Qy      545 S---HADGCHPTTGHCN-----CLPGWSGVHCDSVCAEGRWGPNC---SLPCY 586
```

Db 783 APGFAGPDCRININECQSSPCAFGATCVDEINGYRC--ICPPGRSGPGCQEVGTGRPCF 838

RESULT 10

US-08-185-432-19

; Sequence 19, Application US/08185432

; Patent No. 5750652

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, Spyridon

; APPLICANT: Busseau, Isabelle

; APPLICANT: Diederich, Robert J.

; APPLICANT: Xu, Tian

; APPLICANT: Matsuno, Kenji

; TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND

; TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PENNIE & EDMONDS

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036-2711

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/185,432

; FILING DATE: 21-JAN-1994

; CLASSIFICATION: 530

; ATTORNEY/AGENT INFORMATION:

; NAME: Misrock, S. Leslie

; REGISTRATION NUMBER: 18,872

; REFERENCE/DOCKET NUMBER: 7326-006

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 790-9090

; TELEFAX: (212) 869-8864/9741

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 19:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2703 amino acids

; TYPE: amino acid

; TOPOLOGY: unknown

; MOLECULE TYPE: protein

US-08-185-432-19

Query Match 18.8%; Score 676; DB 1; Length 2703;

Best Local Similarity 25.4%; Pred. No. 1.9e-35;

Matches 208; Conservative 79; Mismatches 202; Indels 330; Gaps 51;

Qy 7 SCL---SFICLLLCHWIGTASPLNLED--PNVCSHWESYSVTVQESYPHPFDQIYYTSC 60

||| :| | : : || : : : | | : : |

Db 502 SCLDDPGTFRVCMPGFTGTQCEIDIDECQSNPC-----LNDGTC 541

Qy 61 TDILNWFKCTRHVSRTAYRHGEKTMYYRRKSQCCPGFYESGEMC---VPHCADKCVHGR 117
| : | | | : | | : | : |
Db 542 HDKINGFKCS-----CALGF--TGARCQINIDDCQSQPCNR 576

Qy 118 -----CIAPNTCQCEPGWGGTNCs---SACDGDHWGPHCTSRCQCKNGALCNPITG-ACH 168
| | : | | | : | | : | | : |
Db 577 GICHDSIAGYSCECPPGYTGTSC EININDCDSN-----PCHRGKCIDDVNSFKCL 626

Qy 169 CAAGFRGW-----RCEDR-----CEQGTYG-----NDCHQRCQ 196
| | : | : | : | | | : | |
Db 627 CDPGYTGYICQKQINECESNPCQFDGHCQDRVGSYYCQCQAGTSGKNCEVNVNECHSN-P 685

Qy 197 CQNGATC-DHVTG-ECRCPPGYTGAFCELDLCPGKHGPQCEQRCPCQNGGVC-HHVTG-E 252
| | | | | | : : | | | | | | : : | | | | | :
Db 686 CNNGATCIDGINSYKQCQVPGFTGQHCE-----KNVDECIS-SPCANNGVCIDQVNGYK 738

Qy 253 CSCPSGWMGTVC-----GQP-----CPEGRFGKNCS---QECQ-- 282
| | | : | | | | | | | |
Db 739 CECPRGFYDAHCLSDVDECASNPCVNEGRCEDGINEFICHCPPGYTGKRCELDIDECCSN 798

Qy 283 -CHNGGTC-DAATG-QCHCSPGYTGERCQ---DECPVGTYGVLCAETCQCVNGGKCY-HV 335
| : | | | | | | | | | | : | : | | | | |
Db 799 PCQHGGTCYDKLNAFSCQCMPPGYTGQKCE TNIDDC-----VTNPCGNGGT CIDKV 848

Qy 336 SG-ACLCEAGFAGERCEARLCPEGLYGIKCDK-RCPCHLENTHSCHPMSG-----ECACKP 389
: | : | : | | | | : | : | : | | | | |
Db 849 NGYKCVCKVPFTGRDCESKMDP-----CARNRC---KNEAKCTPSSNFLDFSTCKL 897

Qy 390 GWSGLYCNE-----TCSPGFYGEAC---QQICS---CQN 417
| : : | | : | | : | | | | : | | |
Db 898 GYTGRYCEDEDIDEC SLSPCRNGASCLNVPGSYRCLCTKGYEGRDCAINTDDCASFP CQN 957

Qy 418 GADCDSVTG--KCTCAPGFKGIDCST-----PCPLGTYGI 450
| | | | | | | | | | | |
Db 958 GRTCLDGIGDYSCLCVDGFDGKH CETDINECLSQPCQNGATCSQYVNSYTCTCPLGFSGI 1017

Qy 451 NCS-----SRCGCKNDVCS PVDG----SCTCKAGWHGVDCSIR----- 485
| | : | | | : | | : | | | : | :
Db 1018 NCQTND EDC TESSCLNGGSC--IDGINGYNC SCLAGYSGANCQYKLNKCD SNPC LINGATC 1075

Qy 486 -----CPSGTWGF GCNL-----TCQCLNGGACNTL--DGTCTCAPGWRGEKCE- 526
| | | | | | : | | | : : : | : | | : | :
Db 1076 HEQNN EYTCHCPSGFTGKQCSEYVDWCGQSPCENGATCSQMKHQF SCKCSAGWTGKLC DV 1135

Qy 527 --LPCQDGT--YGLNCAERCD---CSHADGCHPTTGHCRC L PGWSGVHC----- 568
: | | | | : : | : | | | | : : | : |
Db 1136 QTISCQDAADRKGLSLRQLCNNGTCKDYGN SHV----CYCSQGYAGSYCQKEIDECQSQP 1191

Qy 569 -----DSVCAEGRWGPNC SL-----PC 585
: | : | | | | |
Db 1192 CQNGGTCDRLIGAYECQCRQGFQGNCELNIDDCAPNPC 1230

RESULT 11

US-08-899-232-4

; Sequence 4, Application US/08899232

; Patent No. 6436650

```
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Qi, Huilin
; TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON
; FILE REFERENCE: 7326-046
; CURRENT APPLICATION NUMBER: US/08/899,232
; CURRENT FILING DATE: 1997-07-23
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 2703
; TYPE: PRT
; ORGANISM: Drosophila sp.
US-08-899-232-4
```

```
Query Match          18.8%; Score 676; DB 4; Length 2703;
Best Local Similarity 25.4%; Pred. No. 1.9e-35;
Matches 208; Conservative 79; Mismatches 202; Indels 330; Gaps 51;
```

```
QY      7 SCL----SFICLLLCHWIGTASPLNLED--PNVCSHWESYSVTVQESYPHPFDQIYYTSC 60
      ||| :| |: : : || :::: | | : :|
Db      502 SCLDDPGTFRFCVCMPGFTGTQCEIDIDECQSNPC-----LNDGTC 541

QY      61 TDILNWFKCTRHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMC---VPHCADKCVHGR 117
      | :| |||: | || :| | : | : |
Db      542 HDKINGFKCS-----CALGF--TGARCQINIDDCQSQPCRN 576

QY      118 -----CIAPNTCQCEPGWGGTNCS---SACDGDHWGPHCTSRCQCKNGALCNPITG-ACH 168
      || :|:| ||: ||:| : || : | | : : |
Db      577 GICHDSIAGYSCECPPGYTGTSCEININDCDSN-----PCHRGKCIDDVNSFKCL 626

QY      169 CAAGFRGW-----RCEDR-----CEQGTYG-----NDCHQRCQ 196
      | |: |: | :|| | : || | | :||
Db      627 CDPGYTG YICQKQINECESNPCQFDGHCQDRVGSYYCQCQAGTSGKNCEVNVNECHSN-P 685

QY      197 CQNGATC-DHVTG-ECRCPPGYTGAFCEDLCPGKGHPQCEQRCPCQNGGVC-HHVTG-E 252
      | |||| | : :|:| ||:| || | : :| || | || | | :
Db      686 CNNGATCIDGINSYKQCQVPGFTGQHCE-----KNVDECIS-SPCANNGVCIDQVNGYK 738

QY      253 CSCPSGWMGTVC-----GQP-----CPEGRFGKNCS---QECQ-- 282
      | || |: | | | | | | | | | | |
Db      739 CECPRGFYDAHCLSDVDECASNPCVNEGRCEDGINEFICHCPPGYTGKRCELDIDECSSN 798

QY      283 -CHNGGTC-DAATG-QCHCSPGYTGERCQ---DECPVGTYGVLCAETCQCVNGGKCY-HV 335
      | :||| | | | ||||:|:| :| | | || | | |
Db      799 PCQHGGTCYDKLNAFSCQCMPPGYTGQKCETNIDDC-----VTNPCGNGGTCIDKV 848

QY      336 SG-ACLCEAGFAGERCEARLCPEGLYGIKCDK-RCPCHLENTHSCHPMMSG----ECACKP 389
      :| |:|: | | ||:| | : || :| | | | | |
Db      849 NGYKCVCKVPFTGRDCESKMDP-----CARNRC---KNEAKCTPSSNFLDFSCTCKL 897

QY      390 GWSGLYCNE-----TCSPGFYGEAC---QQICS---CQN 417
      |:| ||:| | :|:| | | : |||
Db      898 GYTGRYCDDEDIDECSLSSPCRNGASCLNVPGSYRCLCTKGYEGRDCAINTDDCASFPQN 957

QY      418 GADCDSVTG--KCTCAPGFGKIDCST-----PCPLGTYGI 450
      | | | | | | | | | | | | | |
```

Db 958 GRTCLDGIGDYSLCLVDGFDGKHCETDINECLSQPCQNGATCSQYVNSYTCTCPLGFSGI 1017
 Qy 451 NCS-----SRCGCKNDVCSFVDG-----SCTCKAGWHGVDCSIR----- 485
 || : | | | :|| :|:| ||: | :| :
 Db 1018 NCQTNDCTESSCLNGGSC--IDGINGYNCSCLAGYSGANQYKLNKCDNPNCLNGATC 1075
 Qy 486 -----CPSGTWGFGCNL-----TCQCLNGGACNTL--DGTCTCAPGWRGEKCE- 526
 |||| | | : | || | : : :| | : || | : | :
 Db 1076 HEQNNYETCHCPSGFTGKQCSEYVDWCGQSPCENGATCSQMKHQFSCKCSAGWTGKLCDV 1135
 Qy 527 --LPCQDGT--YGLNCAERCD---CSHADGCHPTTGHCRCPLPGWSGVHC----- 568
 : ||| ||: : | : | | | :| :|
 Db 1136 QTISCQDAADRKGLSLRQLCNNGTCKDYGNSHV----CYCSQGYAGSYCQKEIDECQSQP 1191
 Qy 569 -----DSVCAEGRWGPNCSL-----PC 585
 : | :| | || | ||
 Db 1192 CQNGGTCRDLIGAYECQCRQGFQGNCELNIDDCAPNPC 1230

RESULT 12

US-08-185-432-16

; Sequence 16, Application US/08185432

; Patent No. 5750652

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, Spyridon

; APPLICANT: Busseau, Isabelle

; APPLICANT: Diederich, Robert J.

; APPLICANT: Xu, Tian

; APPLICANT: Matsuno, Kenji

; TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND

; TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PENNIE & EDMONDS

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036-2711

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/185,432

; FILING DATE: 21-JAN-1994

; CLASSIFICATION: 530

; ATTORNEY/AGENT INFORMATION:

; NAME: Misrock, S. Leslie

; REGISTRATION NUMBER: 18,872

; REFERENCE/DOCKET NUMBER: 7326-006

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 790-9090

; TELEFAX: (212) 869-8864/9741

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 16:

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2471 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-185-432-16
```

```
Query Match 18.5%; Score 666.5; DB 1; Length 2471;
Best Local Similarity 25.8%; Pred. No. 7.1e-35;
Matches 225; Conservative 74; Mismatches 250; Indels 323; Gaps 57;
```

```
Qy      3 ISLNSCLSFICL----LLCH-----WIGTASPLNLE--DPNVCSHW-----ES 39
      | :: | | | | : | : | | : | | : |
Db      531 IDIDDCSSTPCLNGAKCIDHPNGYECQCATGFTGVLCEENIDNCDPDPCHHGQCQDGIDS 590

Qy      40 YSVTVQESYPHPF--DQI--YYTS-----CTDILNWFKC-----TRHRVSYRTAY 80
      | : | | | | | | : | | : | : | : |
Db      591 YTCICNPGYMGAIKSDQIDECYSSPCLNDGRCIDLVNGYQCNCQPGTSGVNCIENFDDCA 650

Qy      81 R----HG--EKTMYRRKSQCCPGF-----YESGEMCV----- 106
      | | : | | | | | : |
Db      651 SNPCIHGICMDGINRYSCVCSPGFTGQRCNIDIDECASNPCRKGATCINGVNGFRICPE 710

Qy     107 ----PHC-----ADKCVHGRC---IAPNTCQCEPGWGGTNCSSACDGDHWGPHCTSR 151
      | | : | | | | : | | : | | | | | : |
Db     711 GPHHPSCYSQVNECLSNPCIHGNCTGGLSGYKCLCDAGWVGINCE--VDKN----ECLSN 764

Qy     152 CQCKNGALC-NPITG-ACHCAAGFRGWRCE---DRC-----EQGTYGNDCH-QRCQC-- 197
      | : | | | : | | | | : | : | | | : |
Db     765 -PCQNGGTCDNLVNGYRCTCKKGFKGYNQVNIDECASNPCLNQGTCTFDDISGYTCHCVL 823

Qy     198 -QNGATCDHVTGECRCPPGYTGAFCE-----LCPPGKHGPQCE---QRC---PCQ 241
      | | | | | | | : | | | | | : | | |
Db     824 PYTGKNCQTVLAPCSPNPCENAAVKESPNFESYTCLCAPGWQGGQRCCTIDIDECISKPCM 883

Qy     242 NGGVCHHVTGE--CSCPSGWMGTVCQPCPEGRFQKNCQEC---QCHNGGTC--DAATG 294
      | | : | : | | | | : | | : | | | | |
Db     884 NHGLCHNTQGSYMCECPPGFSGMDCEDI-----DDCLANPCQNGGSCMDGVNTF 933

Qy     295 QCHCSPGYTGERCQ---DECPVGTYGVLCAETCQCVNKGK--YHVSAGLCEAGFAGER 349
      | | | : | : | | : | | : | | | | | |
Db     934 SCLCLPGFTGDKCQTMNEC-----LSEPCK--NGGTCSDVNSYTCCKCQAGFDGVH 983

Qy     350 CEARL-----CPEGLYGIKCDKRCP-----CHLENTHSCHPMSGECACK 388
      | | : | : | | | | | | | | | |
Db     984 CENNINECTESSCFNGGTCVDGINSFSC--LCPVGFTGSFCLHEINECSSHPCLNEGTCV 1041

Qy     389 PGWSGLYCNETCSPGFYGEACQ---QICS---CQNGADC--DSVTGKCTCAPGFKGIDCS 440
      | | : | : | | | : | | : | | | : |
Db    1042 DGLGTYRC--SCPLGYTGKNCQTLVNLCSRSPCKNKGTCVQKKAESQCLCPSGWAGAYCD 1099

Qy     441 TP-----CPLGTYGINCSSR---CG---CK 459
      | | | | | | : | | |
Db    1100 VPNVSCDIAASRRGVLEHLCQHSGVCINAGNTHYCQCPLGYTGSYCEEQLDECASNPCQ 1159

Qy     460 NDAVCSPVDGS--CTCKAGWHGVDCSIR-----CPSGTWG 492
      : | | | | | : | | | |
```

Db 1160 HGATCSDFIGGYRCECVPGYQGVNCEYEVEDECQNQPCQNGGTCTDLVNHFKCSCPPGTRG 1219

Qy 493 FGC--NL-TC----QCLNGGACNTLDG--TCTCAPGWRGEKCE-----LPC-QDGTY 534
| | : | | | | | | : | | | : | | : | | : | |

Db 1220 LLCEENIDDCARGPHCLNGGQCMDRIGGYSCRLPGFAGERCEGDINECLSNPCSSEGS- 1278

Qy 535 GLNCAE-----RCDCSHA-----DGC-----HPTTGHCRCPLPGW 563
| : | : | | | | | | | | | | | | | | | |

Db 1279 -LDCIQLTNDYLCVCRSAFTGRHCETFVDVCPQMPCLNGGTCAVASNMPDGFICRCPPGF 1337

Qy 564 SGVHCDSVCAEGRW-----GPNCSLP 584
| | | | : : | | | |

Db 1338 SGARCQSSCGQVKCRKGEQCVHTASGPRCFCP 1369

RESULT 13

US-08-083-590A-19

; Sequence 19, Application US/08083590A

; Patent No. 5786158

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, S. et al.

; TITLE OF INVENTION: Therapeutic And Diagnostic Methods

; TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And

; TITLE OF INVENTION: Nucleic Acids

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/083,590A

; FILING DATE: 25-JUN-1993

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Misrock, S. Leslie

; REGISTRATION NUMBER: 18,872

; REFERENCE/DOCKET NUMBER: 7326-015

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212 790-9090

; TELEFAX: 212 8698864/9741

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 19:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2471 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: unknown

; MOLECULE TYPE: peptide

US-08-083-590A-19

Query Match 18.5%; Score 666.5; DB 1; Length 2471;
 Best Local Similarity 25.8%; Pred. No. 7.1e-35;
 Matches 225; Conservative 74; Mismatches 250; Indels 323; Gaps 57;

Qy	3	ISLNSCLSFCICL-----LLCH-----WIGTASPLNLE--DPNVCSHW-----ES	39
		: : : : :	
Db	531	IDIDDCSSTPCLNGAKCIDHPNGYECQCATGFTGVLCEENIDNCDPDPCHHGQCQDGIDS	590
Qy	40	YSVTVQESYPHPF--DQI--YYTS-----CTDILNWFKC-----TRHRVSRYTAY	80
		: : : : :	
Db	591	YTCICNPGYMGAIKSDQIDECYSSPCLNDGRCIDLVNGYQCNCQPGTSGVNCEINFDDCA	650
Qy	81	R----HG--EKTMYRRKSQCCPGF-----YESGEMCV-----	106
		: :	
Db	651	SNPCIHGICMDGINRYSCVCSPGFTGQRCNIDIDECASNPCRKGATCINGVNGFRCICPE	710
Qy	107	----PHC-----ADKCVHGRC---IAPNTCQCEPGWGGTNCSSACDGDHWGPHCTSR	151
		: : : :	
Db	711	GPHHPSCYSQVNECLSNPCIHGNCCTGGLSGYKCLCDAGWVGINCE--VDKN----ECLSN	764
Qy	152	CQCKNGALC-NPITG-ACHCAAGFRGWRCE---DRC-----EQGTYGNDCH-QRCQC--	197
		: : : :	
Db	765	-PCQNGGTCDNLVNGYRCTCKKGFGKYNQVNIDECASNPCLNQGTCTCFDDISGYTCHCVL	823
Qy	198	-QNGATCDHVTGECRCPPGYTGAFCE-----LCPPGKHGPQCE---QRC---PCQ	241
		: :	
Db	824	PYTGKNCQTVLAPCSNPNCENAAVCKESP NFESYTCLCAPGWQGRCTIDIDECISKPCM	883
Qy	242	NGGVCHHVTGE--CSCPSGWMGTVCQPCPEGRFGKNCSEQE---QCHNGGTC--DAATG	294
		: : : : :	
Db	884	NHGLCHNTQGSYMCECPPGFSGMDCEEDI-----DDCLANPCQNGGSCMDGVNTF	933
Qy	295	QCHCSPGYTGERCQ---DECPVGTYGVLCAETCQCVNKGK--YHVSGACLCEAGFAGER	349
		: : : : : :	
Db	934	SCLCLPGFTGDKCQTDMEC-----LSEPCK--NGGTCSDYVNSYTKCKQAGFDGVH	983
Qy	350	CEARL-----CPEGLYGIKCDKRCP-----CHLENTHSCHPMMSGECACK	388
		: : :	
Db	984	CENNINECTESSCFNGGTCVDGINSFSC--LCPVGFTGSGFCLHEINECSSHPCLNEGTCV	1041
Qy	389	PGWSGLYCNETCSPGFYGEACQ---QICS---CQNGADC--DSVTGKCTCAPGFKGIDCS	440
		: : : : : :	
Db	1042	DGLGTYRC--SCPLGYTGKNCQTLVNLCSRSPCKNKGTVCVQKKAESQCLCPSGWAGAYCD	1099
Qy	441	TP-----CPLGTYGINCSSR---CG---CK	459
		:	
Db	1100	VPNVSCDIAASRRGVLEHLCQHSGVCINAGNTHYCQCPLGYTGSYCEEQLDECASNPCQ	1159
Qy	460	NDAVCSPVDGS--CTCKAGWHGVDCSIR-----CPSGTWG	492
		: : :	
Db	1160	HGATCSDFIGGYRCECVPGYQGVNCEYEVDECQNQPCQNGGTCLDLVNHFCKSCPPGTRG	1219
Qy	493	FGC--NL-TC---QCLNGGACNTLDG--TCTCAPGWRGEKCE-----LPC-QDGTY	534
		: : :	
Db	1220	LLCEENIDDCARGPHCLNGGQCMDRIGGYSCRCLPGFAGERCEGDINECLSNPCSSSEGS-	1278

Qy 535 GLNCAE-----RCDCSHA-----DGC-----HPTTGHCRCLPGW 563
 | : | : | | | | | | | | :
 Db 1279 -LDCIQLTNDYLCVCRSAFTGRHCETFVDVCPQMPCLNNGGTCAVASNMPDGFICRCPGPF 1337

Qy 564 SGVHCDSVCAEGRW-----GPNCSTP 584
 | | | | : : | | | |
 Db 1338 SGARCQSSCGQVKCRKGEQCVHTASGPRCFPC 1369

RESULT 14

US-08-532-384-19

; Sequence 19, Application US/08532384

; Patent No. 6083904

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, S. et al.

; TITLE OF INVENTION: Therapeutic And Diagnostic Methods

; TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And

; TITLE OF INVENTION: Nucleic Acids

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/532,384

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/083,590

; FILING DATE: 25-JUN-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Misrock, S. Leslie

; REGISTRATION NUMBER: 18,872

; REFERENCE/DOCKET NUMBER: 7326-015

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212 790-9090

; TELEFAX: 212 8698864/9741

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 19:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2471 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: unknown

; MOLECULE TYPE: peptide

US-08-532-384-19

Query Match 18.5%; Score 666.5; DB 3; Length 2471;

Best Local Similarity 25.8%; Pred. No. 7.1e-35;

Matches 225; Conservative 74; Mismatches 250; Indels 323; Gaps 57;

```
Qy      3 ISLNSCLSFICL----LLCH-----WIGTASPLNLE--DPNVCSHW-----ES 39
      | :: | | | : | : | : | : | : |
Db      531 IDIDDCSSTPCLNGAKCIDHPNGYECQCATGFTGVLCEENIDNCDPDPCHHGQCQDGDIDS 590

Qy      40 YSVTVQESYPHPF--DQI--YYTS-----CTDILNWFKC-----TRHRVSYRTAY 80
      | : | | | | : | | : | : | : | : |
Db      591 YTCICNPGYMGAIKSDQIDECYSSPCLNDGRCIDLVNGYQCNCQPGTSGVNCEINFDDCA 650

Qy      81 R----HG--EKTMYRRKSQCCPGF-----YESGEMCV----- 106
      | | : | | | | | : |
Db      651 SNPCIHGICMDGINRYSCVCSPGFTGQRCNIDIDECASNPCRKGATCINGVNGFRICICPE 710

Qy     107 ----PHC-----ADKCVHGRC---IAPNTCQCEPGWGGTNCSSACDGDHWGPHCTSR 151
      | | : | : | | : : | | | | | : | |
Db     711 GPHHPSCYSQVNECLSNPCIHGNCCTGGLSGYKCLCDAGWVGINCE--VDKN----ECLSN 764

Qy     152 CQCKNGALC-NPITG-ACHCAAGFRGWRCE---DRC-----EQGTYGNDCH-QRCQC-- 197
      | : | | | : | | | | : | : | | | : | |
Db     765 -PCQNGGTCDNLVNGYRCTCKKGFKGYNQVNIDECASNPCLNQGTCFDDISGYTCHCVL 823

Qy     198 -QNGATCDHVTGECRCPPGYTGAFCED-----LCPPGKHGPQCE---QRC---PCQ 241
      | | | | | | | | : | : | | | | | : | |
Db     824 PYTGKNCQTVLAPCSPNPCENAAVCKESPNEFESYTCLCAPGWQGGQRCRTIDIDEKISKPCM 883

Qy     242 NGGVCHHVTGE--CSCPSGWMGTVCQGQPCPEGRFGKNCSQEC---QCHNGGTC--DAATG 294
      | | : | : | | | | : | | : : | | | : | |
Db     884 NHGLCHNTQGSYMCECPPGFGSGMDCEEDI-----DDCLANPCQNGGSCMDGVNTF 933

Qy     295 QCHCSPGYTGERCQ---DECPVGTYGVLCAETCQCVNNGGKC--YHVSGACLCEAGFAGER 349
      | | | : | : | : | : | | : | | | | | : | : |
Db     934 SCLCLPGFTGDKCQTDMEC-----LSEPCCK--NGGTCSDYVNSYTCCKQAGFDGVH 983

Qy     350 CEARL-----CPEGLYGIKCDKRCP-----CHLENTHSCHPMSGECACK 388
      | | : | : | | | | | | | | | | | |
Db     984 CENNINECTESSCFNGGTCDVDGINSFSC--LCPVGFTGSFCLHEINECSSHPCLNEGTCV 1041

Qy     389 PGWSGLYCNETCSPGFYGEACQ---QICS---CQNGADC--DSVTGKCTCAPGFKGIDCS 440
      | | | : | : | : | : | | : | | | : | | |
Db    1042 DGLGTYRC--SCPLGYTGKNCQTLVNLCSRSPCKNKGTCVQKKAESQCLCPSGWAGAYCD 1099

Qy     441 TP-----CPLGTYGINCSSR---CG---CK 459
      | | | | | | | | : | | | : |
Db    1100 VPNVSCDIAASRRGVLVEHLCQHSGVCINAGNTHYCQCPGTYGTSYCEEQLDECASNPCQ 1159

Qy     460 NDAVCSFVDGS--CTCKAGWHGVDCSIR-----CPSGTWG 492
      : | | | | | : | : | | | | |
Db    1160 HGATCSDFIGGYRCECVPGYQGVNCEYEVDCEQNPQPCQNGGTCIDLNVHFKCSCPPGTRG 1219

Qy     493 FGC--NL-TC---QCLNGGACNTLDG--TCTCAPGWRGEKCE-----LPC-QDGTY 534
      | | : | | | | | | : | | | : | : | | : |
Db    1220 LLCEENIDDCARGPHCLNGGQCMDRIGGYSCRCLPGFAGERCEGDINECLSNPCSSSEGS- 1278

Qy     535 GLNCAE-----RCDCSHA-----DGC-----HPTTGHCRCCLPGW 563
      | : | : | | | | | | | | | | | |
Db    1279 -LDCIQLTNDYLCVCRSAFTGRHCETFDVDCPQMPCLNGGTCAVASNMPDGFICRCPPGF 1337
```


Qy 564 SGVHCDSVCAEGRW-----GPNCSLP 584
|| | | | : : || | |
Db 1338 SGARCQSSCGQVKCRKGEQCVHTASGPRCFCP 1369

RESULT 15

US-08-899-232-1

; Sequence 1, Application US/08899232
; Patent No. 6436650
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Qi, Huilin
; TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON
; FILE REFERENCE: 7326-046
; CURRENT APPLICATION NUMBER: US/08/899,232
; CURRENT FILING DATE: 1997-07-23
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2471
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-899-232-1

Query Match 18.5%; Score 666.5; DB 4; Length 2471;
Best Local Similarity 25.8%; Pred. No. 7.1e-35;
Matches 225; Conservative 74; Mismatches 250; Indels 323; Gaps 57;

Qy 3 ISLNSCLSFICL----LLCH-----WIGTASPLNLE--DPNVCSHW-----ES 39
| :: | | | : | : | : | | : |
Db 531 IDIDDCSSTPCLNGAKCIDHPNGYECQCATGFTGVLCEENIDNCDPDPCHHGQCQDGIDS 590

Qy 40 YSVTVQESYPHPF--DQI--YYTS-----CTDILNWFKC-----TRHRVSYRTAY 80
| : | | | | : | | : : | : : : :
Db 591 YTCICNPGYMGAIKSDQIDECYSSPCLNDGRCIDLVNGYQCNCQPGTSGVNCIEINFDDCA 650

Qy 81 R---HG--EKTMYRRKSQCCPGF-----YESGEMCV----- 106
| | : | | | | :
Db 651 SNPCIHGICMDGINRYSCVCSPGFTGQRCNIDIDECASNPCRKGATCINGVNGFRCICPE 710

Qy 107 ---PHC-----ADKCVHGRC---IAPNTCQCEPGWGGTNCSSACDGDHWGPHCTSR 151
| | : : | | : : | | : | | | : | |
Db 711 GPHHPSCYSQVNECLSNPCIHGNCTGGLSGYKCLCDAGWVGINCE--VDKN---ECLSN 764

Qy 152 CQCKNGALC-NPITG-ACHCAAGFRGWRCE---DRC-----EQGTYGNDCH-QRCQC-- 197
| : | | | : | | | | : : | : | | : | |
Db 765 -PCQNGGTCDNLVNGYRCTCKKGFKGYNCQVNIDECASNPCLNQGTCTFDDISGYTCHCVL 823

Qy 198 -QNGATCDHVTGECRCPPGYTGAFCE-----LCPPGKHGPQCE---QRC---PCQ 241
| | | | | : : | | | | : | | |
Db 824 PYTGKNCQTVLAPCSPNPCENAAVCKESPINFESYTCLCAPGWQGGQRCCTIDIDECISKPCM 883

Qy 242 NGGVCHHVTGE--CSCPSGWMGTVCQPCPEGRFGKNCSQEC---QCHNGGTC--DAATG 294
| | : : | | | : | | : | | | : | |
Db 884 NHGLCHNTQGSYMCECPGFGSGMDCEEDI-----DDCLANPCQNGGSCMDGVNTF 933

Qy 295 QCHCSPGYTGERCQ---DECPVGTYGVLCETCQCVNGGKC--YHVSAGACLCEAGFAGER 349
 | | | | : | : | | : | | : | | | | : | | | |
 Db 934 SCLCLPGFTGDKCQTDMEC-----LSEPCK--NGGTCSDYVNSYTCCKCQAGFDGVH 983

Qy 350 CEARL-----CPEGLYGIKCDKRCP-----CHLENTHSCHPMMSGECACK 388
 | | : | : | | : | | | | | | | | | | | |
 Db 984 CENNINECTESSCFNGGTCVDGINSFSC--LCPVGFTGSFCLHEINECSSHPCLNEGTCV 1041

Qy 389 PGWSGLYCNETCSPGFYGEACQ---QICS---CQNGADC--DSVTGKCTCAPGFKGIDCS 440
 | | : | | : | | : | | : | | : | | : | |
 Db 1042 DGLGTYRC--SCPLGYTGKNCQTLVNLCSRSRSPCKNKGTCVQKKAESQCLCPSGWAGAYCD 1099

Qy 441 TP-----CPLGTYGINCSSR---CG---CK 459
 | | | | | | : | | :
 Db 1100 VPNVSCDIAASRRGVIVEHLCQHSGVCINAGNTHYCQCPLGYTGSYCEEQLDECASNPCQ 1159

Qy 460 NDAVCSPVDGS--CTCKAGWHGVDCSIR-----CPSGTWG 492
 : | | | | | : | | : | | | |
 Db 1160 HGATCSDFIGGYRCECVPGYQGVNCEYEVDECQNQPCQNGGTCIDLNVHFKCSCPPGTRG 1219

Qy 493 FGC--NL-TC----QCLNGGACNTLDG--TCTCAPGWRGEKCE-----LPC-QDGTY 534
 | | : | | | | | : | | | : | | : | | : | |
 Db 1220 LLCEENIDDCARGPHCLNGGQCMDRIGGYSCRCLPGFAGERCEGDINECLSNPCSEGS- 1278

Qy 535 GLNCAE-----RCDCSHA-----DGC-----HPTTGHCRCCLPGW 563
 | : | : | | | | | | | | | | : | | : | |
 Db 1279 -LDCIQLTNDYLCVCRSAFTGRHCETFDVCPQMPCLNGGTCAVASNMPDGFICRCPPEF 1337

Qy 564 SGVHCDSVCAEGRW-----GPNCSLP 584
 | | | | : : | | | |
 Db 1338 SGARCQSSCGQVKCRKGEQCVHTASGPRCFPC 1369

Search completed: March 26, 2004, 16:13:10
 Job time : 18.241 secs

OM protein - protein search, using sw model

Run on: March 26, 2004, 16:05:25 ; Search time 11.883 Seconds
(without alignments)
4743.616 Million cell updates/sec

Title: US-10-092-390-4
Perfect score: 3601
Sequence: 1 MVISLNSCLSFI~~CL~~LLCHWI.....HCDSVCAEGRWGPNC~~SL~~PCY 586

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query	Match Length			
1	1372.5	38.1	1574	2	T13954	MEGF6 protein - ra
2	1322	36.7	1620	2	T27283	hypothetical prote
3	1284.5	35.7	1111	2	T26972	hypothetical prote
4	719	20.0	2524	2	A35844	Xotch protein - Af
5	717	19.9	2437	2	S42612	transmembrane prot
6	697	19.4	2321	2	S78549	notch3 protein - h
7	693	19.2	2531	2	A46019	notch-1 protein -
8	687	19.1	1203	2	A49175	Motch B protein -
9	685.5	19.0	2318	2	S45306	notch 3 protein -
10	685.5	19.0	2471	2	A49128	cell-fate determin
11	682	18.9	3566	1	A40701	tenascin-X precurs
12	681.5	18.9	2531	2	T31070	notch homolog - se
13	677	18.8	2703	1	A24420	notch protein - fr
14	676.5	18.8	4006	2	T09070	probable tenascin

15	675	18.7	2531	2	S18188	notch protein homo
16	672.5	18.7	4135	2	T42629	tenascin-X - bovin
17	664.5	18.5	1064	2	A40136	fibropellin Ia - s
18	659.5	18.3	2555	2	A40043	notch protein homo
19	658	18.3	1964	2	T09059	notch4 - mouse
20	648.5	18.0	2352	2	T30201	Notch homolog prot
21	644.5	17.9	2201	2	A32160	tenascin-C - human
22	640.5	17.8	2019	1	JQ1322	tenascin precursor
23	636	17.7	1810	1	A32230	tenascin precursor
24	631.5	17.5	2139	2	A35672	crumbs protein - f
25	616	17.1	1220	2	A56136	jagged protein pre
26	611	17.0	1746	1	S19694	tenascin precursor
27	593.5	16.5	3672	2	T23433	hypothetical prote
28	593.5	16.5	3704	2	T37316	probable laminin a
29	587	16.3	1408	2	S16148	gene serrate prote
30	586	16.3	1722	2	E89753	protein Fl1C7.4 [i
31	586	16.3	3635	2	T10053	laminin alpha 5 ch
32	580.5	16.1	861	2	A48825	Notch homolog Motc
33	577.5	16.0	1801	1	MMRTS	laminin beta-2 cha
34	576.5	16.0	3712	2	S18253	laminin alpha-1 ch
35	565.5	15.7	2823	2	T23064	hypothetical prote
36	565.5	15.7	2823	2	F87908	protein T22A3.8 [i
37	565.5	15.7	3102	2	T43291	laminin alpha chai
38	562.5	15.6	647	2	A43902	tenascin - eastern
39	561	15.6	1798	2	S53869	laminin beta-2 cha
40	561	15.6	3106	1	S53868	laminin alpha-2 ch
41	560.5	15.6	1429	2	S06434	homeotic protein l
42	556	15.4	833	2	S19087	gene Delta protein
43	552	15.3	832	2	A31246	neurogenic protein
44	552	15.3	880	2	S00670	neurogenic repetit
45	550	15.3	1797	2	A55677	laminin beta-2 cha

ALIGNMENTS

RESULT 1

T13954

MEGF6 protein - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000

C;Accession: T13954

R;Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.

Genomics 51, 27-34, 1998

A;Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs by motif-trap screening.

A;Reference number: Z14126; MUID:98360089; PMID:9693030

A;Accession: T13954

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-1574 <NAK>

A;Cross-references: EMBL:AB011532; NID:g3449293; PIDN:BAA32462.1; PID:g3449294

A;Experimental source: strain Sprague-Dawley; brain

C;Genetics:

A;Gene: MEGF6

Query Match

38.1%; Score 1372.5; DB 2; Length 1574;

Best Local Similarity 42.2%; Pred. No. 1.6e-68;
Matches 230; Conservative 60; Mismatches 199; Indels 56; Gaps 11;

```

Qy      89 RRKSQCCPGFYESGEMCVPHCADKCVH--GRCIAPNT---CQCEPGWGGTNCSSACDGDHWG 145
      | : | |:| :| | || | :| | | | ||| | :| ||| |||
Db      816 RCQDTCAGWYGTG--CQIRCA--CANDGHC-DPTTGRCSCAPGWTGLSCQQRACDSGHWG 870

Qy      146 PHCTSRCQCKNG-ALCNPITGACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCCQNGATCD 204
      | | | | | | | : :| | | ||: | || | | | | | | :| :| :| :| | |
Db      871 PDCIHPCNCSAGHGNCDAVSGLCLCEAGYEGPRCEQSCRQGYGPSCEQKCRCEHGAACD 930

Qy      205 HVTGECRCPPGYTGAFCD-----LCPPGKHGPQC 234
      ||:| | || | :| :||| :|| | :| ||:|
Db      931 HVSGACTCPAGWRGSFCEHACAPAGFFGLDCDSACNCSAGAPCDAVTGSCICPAGRWGPRC 990

Qy      235 EQRCP-----CQNGGVCHHVTGECSCPSGWMGTVCQGPCPEGRFGKNCSQEC 281
      | || | || | |||:| | ||| | | ||| | :||| |
Db      991 AQSCPPLTFGLNCSQICTCFNGASCDSVTGQCHCAPGWMGPTCLQACPPGLYGKNCQHSC 1050

Qy      282 QCHNGGTCDAAATGQCHCSPGYTGERCQDECPVGTYGVLCAETCQCVNNGGKCYHVSAGCLC 341
      | ||| || ||| | | :|| | :||| | | | | :||| | :| |||
Db      1051 LCRNGGRCDPIIGQCTCPEGWTGLACENECLPGHYAAGCQLNCSCLHGGICDRLTGHCCLC 1110

Qy      342 EAGFAGERCEARLCPEGLYGIKCDKRCPCHELENTHSCHPMSGECACKPGWSGLYCNCTCS 401
      ||:| :| :| :| | | :| :| :| | | ||| :| | ||| | :| :|
Db      1111 PAGWTGDKCQSS-CVSGTFGVHCEEHCAC--RKGASCHHVTGACFCPPGWRGPHCEQACP 1167

Qy      402 PGFYGEACQQICSCQNGADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKND 461
      | :| ||| | | | | |||:| | ||| | :| | | :| | | :
Db      1168 RGWFGEACAQRCLCPTNASCHHVTGECRCPPGFTGLSCEQACQPGTFGKDCEHLCQCPGE 1227

Qy      462 A-VCSPVDGSCTCKAGWHGVDCSIRCPSGTWGFGCNLTCQCLNGGACNTLDGTCTCAPGW 520
      | | | ||| ||:| | |||| :| | | :||| | :| | | | :
Db      1228 TWACDPASGVCTCAAGYHGTGCLQRCPSGRYGPGEHICKCLNGGTCDPATGACYCPAGF 1287

Qy      521 RGEKCELPQDGTGYNCAERCDCSHADGCHPTTGHCRCPLGWSGVHCDVCAEGRWGPN 580
      | | | | | :| :|| | | | :| | | | :|| | :| | :| :|
Db      1288 LGADCSLACPQGRFGPSCAHVCACRQGAACDPVSGACICSPGKTGVRCEHGCPQDRFGKG 1347

Qy      581 CSLPC 585
      | | |
Db      1348 CELKC 1352

```

RESULT 2

T27283

hypothetical protein Y64G10A.f - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C;Accession: T27283

R;Ainscough, R.

submitted to the EMBL Data Library, September 1999

A;Reference number: Z20336

A;Accession: T27283

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-1620 <WIL>

A;Cross-references: EMBL:AL110498; NID:e1542303; PIDN:CAB54471.1; CESP:Y64G10A.f
 A;Experimental source: clone Y64G10A
 C;Genetics:
 A;Gene: CESP:Y64G10A.f
 A;Introns: 77/1; 116/1; 198/1; 282/1; 365/1; 425/1; 466/1; 548/1; 559/1; 601/1;
 625/1; 715/1; 782/3; 845/1; 895/2; 956/1; 1105/1; 1221/1; 1307/1; 1445/2

Query Match 36.7%; Score 1322; DB 2; Length 1620;
 Best Local Similarity 36.3%; Pred. No. 9.9e-66;
 Matches 235; Conservative 76; Mismatches 234; Indels 102; Gaps 12;

```

Qy      16 LCHWI-GTASPLNLEDPNVCSHW-----ESYSVTVQESYPHPFDQIYYTSCDIL 64
      :|| : || : | : :| :||: | | : : :||:
Db      872 VCHHVTGTCTCLPGKTGPLCDQCLIFVETIEFDIAFSINVIACAPNTYGPNCATCS-CV 930

Qy      65 NWFKCTRHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCAD-----KC 113
      | || | | | | | | | | | | | | | |
Db      931 NGAKCDES DGS-----CHCTPGFY--GATCSEVCPTGRFGIDCMQLCKC 972

Qy      114 VHGR-CIAPN-TCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAA 171
      :| | | :|| | | | | :| | :| | :| | :| |
Db      973 QNGAICDTSNGSCECAPGWSGKKCDKACAPGTFGKDCSKKCDCADGMHCDPSDGEICPP 1032

Qy      172 GFRGWRCEDRCEQGTYGNDCHQRCQCGNGAT----- 202
      | :| :||: | :| | | | | | | |
Db      1033 GKKGHKCDETCDSGLFGAGCKGICSCQNGATCDSVTGSCECRPGWRGKKCDRPCPDGRFG 1092

Qy      203 -----CDHVTGECRCPPGYTGAFCEDLCPGKHGPQCEQRC 238
      | | | | | | | | | | :|| | : | | :|| | |
Db      1093 EGCNAICDCTTTNDTSMYNPFVARCDHVTGECRCPAGWTGPDCQTSCLGRHGEGRHSC 1152

Qy      239 PCQNGGVCHHVTGECSCPSGWMGTVCQGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHC 298
      | || | | | | | | | | | | | :| | :| | :| | :| |
Db      1153 QCSNGASCDRVTFGCDPCPSGFMGNCESECEGLWGSNCMKHCLCMHGGECKENGDCEC 1212

Qy      299 SPGYTGERCQDECPVGTYGVLCAETCQCVNGGKCYHVS GACLCEAGFAGERCEARLCPEG 358
      | :|| | | | :| | | | | | | :| | | | :| | |
Db      1213 IDGWTGPSL---CPFGQFGRNCAQRCNCKNGASCDRKTGRCECLPGWSGEHCE-KSCVSG 1268

Qy      359 LYGIKCDKRCPCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNG 418
      || || :| | | | | | | | | | | | | | | | | |
Db      1269 HYGAKCEETCEC--ENGALCDPISGHCSCQPGWRGKKCNRPCLKGYFGRHCSQSCRCANS 1326

Qy      419 ADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSVPDGSCTCKAGWH 478
      || :||:| | | :| | | | | | | | | :|| :| | | |
Db      1327 KSCDHISGRQCQPKGYAGHSCTELCPDGTFGESCSQKCDGENSEMCDALSGKCFCKPGHS 1386

Qy      479 GVDCSI RCPSGTWGFGCNLTQCCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTGYGLNC 538
      | || | | :| | | | | | | | | | | | | | | |
Db      1387 GSDCKSGCVQGRFGPD CNQLCSCENGGVCDSSSTGSCVCPPGYIGTKCEIACQSDRFGPTC 1446

Qy      539 AERCDCSHADGCHPTTGHCRCLPGWSGVHCDSVCAEGRWGPNCSLPC 585
      : | :| : | | | | | | :| :| | | | | |
Db      1447 EKICNCEGGTCDRLTGQCRCLPGFTGMT'CNQVCPEGRFAGCKEKC 1493

```

RESULT 3

T26972
hypothetical protein Y47H9C.4 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C;Accession: T26972
R;Harris, B.
submitted to the EMBL Data Library, October 1998
A;Reference number: Z20293
A;Accession: T26972
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1111 <WIL>
A;Cross-references: EMBL:AL032657; PIDN:CAA21739.1; GSPDB:GN00019; CESP:Y47H9C.4
A;Experimental source: clone Y47H9C
C;Genetics:
A;Gene: CESP:Y47H9C.4
A;Map position: 1
A;Introns: 50/2; 84/2; 150/1; 238/3; 342/3; 797/1; 851/1; 947/2; 1017/1; 1083/1
C;Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology

Query Match 35.7%; Score 1284.5; DB 2; Length 1111;
Best Local Similarity 34.2%; Pred. No. 9e-64;
Matches 246; Conservative 77; Mismatches 221; Indels 175; Gaps 20;

```

Qy      21 GTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYT-----SCTDILNWFKCTRHR 73
      || | : :||: ||: | : : | | : | | :|
Db      35 GTTEP---QGDHVCT-----VKTIVDDY--ELKKVIHTVVYNDTEQCLNPLTGFQC---- 80

Qy      74 VSYRTAYRHGEKTMYYRK-----SQCCPGFYESGE-MCVPHCADKCVHGRCIAPNTC 124
      | : |:| |:|: ||| |:|: : |:| | | |:| | |
Db      81 ----TVEKRGGKASYQRQLVKKEKYVKQCCDGYQYTKDHFCLPDCNPPCKKGGKCIIEPGKC 136

Qy     125 QCEPGWGGTNCSSACDGDHWGPHCTSRQCCKNGALCNPITGACHCAAGFRGWRCE----- 179
      :|:|:| | |:|: | | |: | |:| | | | |:|:| | |
Db     137 ECDPGYGGKYCASSCSVGTWGLGCSKSCDCENGANCDPELGTCICTSGFQGERCEKPCPD 196

Qy     180 -----DRCEQGTGYNDCHQRCQCQNGAT 202
      :|:|:| |:| :| :| | | | | |
Db     197 NKWGPNCVKSCPCQNGGKCNKEGKVCVSDGWGGEFCLNKCEEGKFGECKFECNCQNGAT 256

Qy     203 CDHVTGECRCPPGYTGAFCEDLCPGKHGPQCEQRCPCQNGGVCHHVTGECSCPSGWMGT 262
      ||: |:| | || || |:| | | | |:| | | | :|| | || |
Db     257 CDNTNGKCICKSGYHGALCENECSVGFFGSGCTQKCDCLNNQNCDSSSGECKC-IGWTGK 315

Qy     263 VCGQPCPEGRFGKNCSQECQC-----HNGGTCAATGQCHCSPGYTGERCQD-ECPVGT 315
      | | | | | | | : :||| ||| | | | |:| : :|
Db     316 HCDIGCSRGRFGLQCKQNCTCPGLEFSDSNASCDAKTGQCQCESGYKGPCKDERKCDAEQ 375

Qy     316 YGVLCAETCQCV--NNGKCYHVSAGACLCEAGFAGERCEARLCPEGLYGIKCDKRCPCHLE 373
      || |:| | | | | :| | |:| | | | : | | |:|: |
Db     376 YGADCSKTCTCVRENTLMCAPNTGFCRCKPGFYGDNCEL-ACSKDSYGPNCEKQAMCDWN 434

Qy     374 NTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSC-QNGADCDSVTGKCTCAP 432
      : |:| :| | ||| :| |:| | || | | | | || | |
Db     435 HASECNPETGSCVCKPGRTGKNCEPCPLDFYGPNCACHQCQCNCQRGVCGDGADGKQCQCDR 494

```

```

Qy      433 GFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSPVDGSCTCKAGWHGVDCSIRCPSGTWG 492
          |: | | | | |: | | | | | |: | | | | |: | | | | |: |
Db      495 GWTGHRCEHHCPADTFGANCKRCKCPKGIGCDPITGECTCPAGLQGANCDIGCPEGSYG 554

Qy      493 FGCNLTQCQLNGGACNTLDGTCTCAPGW-----RGEKCEL--PCQD----- 531
          || | |:|:| | |: | | | | |: | | | | | | |
Db      555 PGCKLHCKCVN-GKCDKETGECTCQPGFFGSDCSTTCSKGKYGESCELSGPCSDASCSKQ 613

Qy      532 -----GTYGLNCAERCD----- 543
          || |:| |:|
Db      614 TGKCLCPLGTKGVSCDQKCDPNTFGFLCQETVTPSPCASTDPKNGVCLSCPPGSSGIHCE 673

Qy      544 -----CSHAD--GCHPTTGHCRLPGWSGVHCDVCAEGRWGPNCSLPC 585
          || | | | | | | | | |: | | | |:|:| |:| |
Db      674 HNCPAGSYGDGCQQVCSCADGHGCDPTTGECICEPGYHGKTCSEKCPDGKYGYGCALDC 732

```

RESULT 4

A35844

Xotch protein - African clawed frog

C;Species: *Xenopus laevis* (African clawed frog)

C;Date: 12-Oct-1990 #sequence_revision 12-Oct-1990 #text_change 02-Aug-2002

C;Accession: A35844

R;Coffman, C.; Harris, W.; Kintner, C.

Science 249, 1438-1441, 1990

A;Title: Xotch, the *Xenopus* homolog of *Drosophila* notch.

A;Reference number: A35844; MUID:90385285; PMID:2402639

A;Accession: A35844

A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-2524 <COF>

C;Superfamily: notch protein; ankyrin repeat homology; EGF homology

C;Keywords: transmembrane protein

F;146-177/Domain: EGF homology <EGX1>

F;184-215/Domain: EGF homology <EGF1>

F;222-254/Domain: EGF homology <EGF>

F;456-487/Domain: EGF homology <EGX2>

F;757-788/Domain: EGF homology <EGF3>

F;1025-1056/Domain: EGF homology <EGX3>

F;1924-1956/Domain: ankyrin repeat homology <AN1>

F;1957-1989/Domain: ankyrin repeat homology <AN2>

F;1991-2023/Domain: ankyrin repeat homology <AN3>

F;2024-2056/Domain: ankyrin repeat homology <AN4>

F;2057-2089/Domain: ankyrin repeat homology <AN5>

Query Match 20.0%; Score 719; DB 2; Length 2524;

Best Local Similarity 25.9%; Pred. No. 1.8e-32;

Matches 225; Conservative 60; Mismatches 221; Indels 364; Gaps 51;

```

Qy      5 LNSCLSFICL-----LLCHWIGTASPLNLED-----PNVCSHW-----ESYS 41
          :| ||| || | | : | : | | | : | : |
Db      604 INECLSKPCLNGGQCTDRENGYICTCPKGTGVNCE TKIDDCASNLCNDNGKCIDKIDGYE 663

Qy      42 VTVQESYPHPFDQIYYT-----SCTDIILNWFKCTRHRVSYRTAYRHGEKTM YRR 90
          | : | | | | : | | | | |
Db      664 CTCEPGYTGKLCNININECDNPCRNGGTCKDQINGFTCV----- 703

```


Qy 91 KSQCCPGFYESGEMC---VPHC-ADKCVHGRC---IAPNTCQCEPGWGGTNC---SSACD 140
 Db 704 ----CPDGYHD-HMCLSEVNECNSNPCIHGACHDGVNGYKCDCEAGWSGSNCDINNNECE 758

Qy 141 GDHWGPHCTSRCQCKNGALCNPTGA--CHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCQ 198
 Db 759 SN-----PCMNGGTCKDMTGAYICTCKAGFSGPNCQ-----TNINECSSN-PCL 801

Qy 199 NGATC-DHVTG-ECRCPPGYTGAFCEDLCPGKHGPQCEQRC---PCQNGGVCHH----V 249
 Db 802 NHGTCIDDVAGYKCNCLPYTGAICEAVLAP-----CAGSPCKNGGRCKESED FE 851

Qy 250 TGECSCPSGWMGTVCQPCPEGRFGKNCSQEC---QCHNGGTCDAAATG---QCHCSPGYTG 304
 Db 852 TFSCECPPGWQGGTC-----EIDMNECVNRPCRNATCQNTNGSYKCNCKPGYTG 901

Qy 305 ERCQ---DECPVGTYGVLCAETCQCVNNGGKCYHVS GA--CLCEAGFAGERCEARL----- 354
 Db 902 RNCEMDIDDC-----QPNPCHNGGSCSDGINMFFCNC PAGFRGPKCEEDINECAS 951

Qy 355 -----CPEGLYGIKCDKRCPC HLE-----NTHSCHPMSG-----ECAC 387
 Db 952 NPCKNGANCTDCVNSYTCTCQPGFSGIHCESNTPDCTESSCFNGGTC--IDGINTFTCQC 1009

Qy 388 KPGWSGLYC---NE-----TCSPGFYGEACQOI---CS---CQ 416
 Db 1010 PPGFTGSYCQHDINECDSPCLNGGTCQDSYGYTKCTCPQGYTGLNCQNLVRWCDSSPCK 1069

Qy 417 NGADCDSVTG--KCTCAPGFKGIDCSTP----- 442
 Db 1070 NGGKCWQTNNFYRCECKSGWTGVYCDVPSVSCEVAAKQQGV DIVHLCRNSGMCVDTGNTH 1129

Qy 443 ---CPLGTYGINCSSR---CG---CKNDAVCSPVDG--SCTCKAGWHGVDCS----- 483
 Db 1130 FCRCQAGYTGSYCEEQVDECSNPNCQNGATCTDYLGGYSCECVAGYHGVNCSSEEINECLS 1189

Qy 484 -----IRCPSGTWGFGCNLT---C-----QCLNGGACNTLDG 512
 Db 1190 HPCQNGGTCIDLINTYKCS CPRGTQGVHCEINVDDCTPFYDSFTLEPKCFNNGKCIDRVG 1249

Qy 513 --TCTCAPGWRGEKCE-----LPCQD-GTYGLNCAE-----RCDC-----SH 546
 Db 1250 GYNICPPPGFVGERCEGDVNECLSNPCDSRG TQ--NCIQLVNDYRCECRQGFTGRRCESV 1307

Qy 547 ADGC-----HPTTGH-CRCLPGWSGVHCD----- 569
 Db 1308 VDGCKGMPCRNGGTCAVASNTERGFICKCPPGFDGATCEYDSRTCSNLRQNGGT CISVL 1367

Qy 570 ----SVCAEGRWGPNC-----SLPCY 586
 Db 1368 TSSKVCVSEGYTGATCQYPVISPCASHPCY 1397

RESULT 5

S42612

transmembrane protein precursor - zebra fish

C;Species: Brachydanio rerio (zebra fish)
 C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 02-Aug-2002
 C;Accession: S42612
 R;Bierkamp, C.; Campos-Ortega, J.A.
 Mech. Dev. 43, 87-100, 1993
 A;Title: A zebrafish homologue of the Drosophila neurogenic gene Notch and its
 pattern of transcription during early embryogenesis.
 A;Reference number: S42612; MUID:94128602; PMID:8297791
 A;Accession: S42612
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-2437 <BIE>
 A;Cross-references: EMBL:X69088; NID:g433866; PIDN:CAA48831.1; PID:g433867
 C;Superfamily: notch protein; ankyrin repeat homology; EGF homology
 F;755-786/Domain: EGF homology <EGF1>
 F;1023-1054/Domain: EGF homology <EGF>
 F;1185-1216/Domain: EGF homology <EGF2>
 F;1915-1947/Domain: ankyrin repeat homology <AN1>
 F;1948-1980/Domain: ankyrin repeat homology <AN2>
 F;1982-2014/Domain: ankyrin repeat homology <AN3>
 F;2015-2047/Domain: ankyrin repeat homology <AN4>
 F;2048-2080/Domain: ankyrin repeat homology <AN5>

Query Match 19.9%; Score 717; DB 2; Length 2437;
 Best Local Similarity 25.7%; Pred. No. 2.2e-32;
 Matches 221; Conservative 60; Mismatches 208; Indels 370; Gaps 49;

```

Qy      4 SLNSCLS-----FICLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESY 48
      ::| |||          :||      |      :|::|      |
Db      601 NINECLSQPCRNGGTCQDRENAYICTCPKGTGVNCEINIDD---CKR----- 645

Qy      49 PHFPDQIYYTSCTDILNWFKCTRHRVSRYRTAYRHGEKTMYYRRKSQCCPGFYESGEMC--- 105
      | |  |  | | :| ::|          | ||:  || ||
Db      646 -KPCD---YGKCIDKINGYECV-----CEPGY--SGSMCNIN 676

Qy      106 VPHCA-----DKCVHGRC---IAPNT 123
      :  ||          : |:||  |  |
Db      677 IDDCALNPCHNGGTCIDGVNSFTCLCPDGFRDATCLSQHNECSSNPCIHGSCLDQINSYR 736

Qy      124 CQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGA--CHCAAGFRGWRCEDR 181
      | || ||  | ||  :      | |  | ||  | :|  | || ||  | :
Db      737 CVCEAGWMGRNCDININ-----ECLSN-PCVNGGTCKDMTSGYLCTCRAGFSGPNCQMN 789

Qy      182 CEQGTYGNDCHQRCQCQNGATC-DHVTG-ECRCPPGYTGAFCEDLCPPGKHGPQCEQRCF 239
      | :|  | |  :| | | | :| |  |||  ||::  |  | | |
Db      790 I-----NECASN-PCLNQGSCIDDVAGFKCNCMLPYTGVCENVLAP-----CSPR-P 835

Qy      240 CQNGGVCHH----VTGECSCPSGWMGTVCQGPCPEGRFGKNCSQEC---QCHNGGTCDAA 292
      |:|||||  :  |:|:| | |  |  ||  | |||  |:
Db      836 CKNGGVCRESEDFQSFSCNPAGWQGQTCCEVDI-----NECVRNPCTNGGVCENL 885

Qy      293 TG--QCHCSPGYTGERCQ---DECPVGTYGVLCAETCQCVNGGKCY-HVSG-ACLCEAGF 345
      |  || |:|:| |  |:  |:|  |  | || |  |:|  |:| |||
Db      886 RGGFQCRCPGFTGALCENDIDDC-----EPNPCSNGGVCQDRVNGFVCVCLAGF 935

Qy      346 AGERCEARL-----CPEGLYGIKCDKRCPCHELENTHSCHP- 380
      ||||  :          || |  || |:  |  |:  ||
  
```

Db 936 RGERCAEDIDECVSAPCRNGGNC TDCVNSYTCSCPAGFSGINCEINTPDCTES--SCFNG 993
 Qy 381 -----MSGECACKPGWSGLYC----NE-----TCSPGFYGEA 408
 | | | | : | | | | | | | : | |
 Db 994 GTCVDGISSFSCVCLPGFTGNYCQHDVNECD SRPCQNGGSCQDGYGTYKCTCPHGYTGLN 1053
 Qy 409 CQQI---CS---CQNGADC--DSVTGKCTCAPGFKGIDCSTP----- 442
 | | : | | | : | | | : | | | : | | | |
 Db 1054 CQSLVRWCDSSPCKNGGSCWQQGASFTCQCASGWTGIYCDVPSVSCEVAARQQGVSVAVL 1113
 Qy 443 -----CPLGTYGINCSSRCG-----CKNDAVCSPVDG--SCTCKAGW 477
 | | | | : | | | : | | | | :
 Db 1114 CRHAGQCVDAGNTHLCRCQAGYTGSYCQEQVDECQPNPCQNGATCTDYLGGYSCECVPGY 1173
 Qy 478 HGVDCS-----IRCPSGTWGFGCNL---TC----- 499
 | | : | | | | | | | : | |
 Db 1174 HGMNCSKEINECLSQPCQNGGTCIDL VNTYKCS CPRGTQGVHCEIDIDDCSPSVDPLTGE 1233
 Qy 500 -QCLNGGACNTLDG--TCTCAPGWRGEKCE-----LPCQ-DGTYGLNCAE-----RC 542
 : | | | | | | | | : | | : | | | | : | |
 Db 1234 PRCFNGGRCVDRVGGYGCVCPAGFVGERCEGDVNECLSDPCDPGSGSY--NCVQLINDFRC 1291
 Qy 543 DCSHA-----DGCHPT-----TGH---CRCLPGWSGVHCD----- 569
 : | | | | : | | | | | | | : | | | | :
 Db 1292 ECRTGYTGKRCETVFENGCKDTPCKNGGTCAVASNTKHGYICKCQPGYSGSSCEYDSQSCG 1351
 Qy 570 -----SVCAEGRWGPNC 581
 : | | | | |
 Db 1352 SLRCRNGATCVSGHLSPRC 1370

RESULT 6

S78549

notch3 protein - human

C;Species: Homo sapiens (man)

C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 08-Sep-2002

C;Accession: S78549; S71825

R;Joutel, A.; Tournier-Lasserre, E.

submitted to the EMBL Data Library, April 1997

A;Reference number: S78549

A;Accession: S78549

A;Molecule type: mRNA

A;Residues: 1-2321 <JOU1>

A;Cross-references: EMBL:U97669; NID:g2668591; PIDN:AAB91371.1; PID:g2668592

R;Joutel, A.; Corpechot, C.; Ducros, A.; Vahedi, K.; Chabriat, H.; Mouton, P.;

Alamowitch, S.; Domenga, V.; Cecillion, M.; Marechal, E.; Maciazek, J.;

Vayssiere, C.; Cruaud, C.; Cabanis, E.A.; Ruchoux, M.M.; Weissenbach, J.; Bach,

J.F.; Bousser, M.G.; Tournier-Lasserre, E.

Nature 383, 707-710, 1996

A;Title: Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia.

A;Reference number: S71825; MUID:97032728; PMID:8878478

A;Accession: S71825

A;Status: nucleic acid sequence not shown

A;Molecule type: DNA

A;Residues: 67-113;138-194;268-333,'G',335-346;536-613;716-765;1240-1279;1815-1888 <JOU2>


```

Qy      353 R-----LCPEGLYGIKCDKRCPCCHLENTHSCHPM----- 381
          |           || | | | : | | |
Db      886 RDVDECLSNPCGPGTCTDHVASFTCTCPPGYGGFHCEQDLP-----DCSPSSCFNGGT 938

Qy      382 -----SGEACKPGWWSGLYCNE-----TCSPGFYGEACQ 410
          | | | : | : | : | | | | |
Db      939 CVDGVNSFSCLCRPGYTGAHCQHEADPCLSRPCLHGGVCSAAHPGFRCTCLESFTGPQCQ 998

Qy      411 QI---CS---CQNGADCDSVTGKCTCAPGFKGIDC---STP----- 442
          : || || | | | | : | | | |
Db      999 TLVDWCSRQPCQNGGRCVQTGAYCLCPPGWSGRLCDIRSLPCREAAAQIGVRLEQLCQAG 1058

Qy      443 -----CPLGTYGINCSSRCG-----CKNDVAVCSPVDGS---CTCKAGWHGVD 481
          || | | : | | : | | | : | :
Db      1059 GQCVDEDSSHYCVCPEGRTGSHCEQEVDPCLAQPCQHGGTTCRGYMGGYMCECLPGYNGDN 1118

Qy      482 CS-----IRCPSGTWGFGCNLT---C-----QCLN 503
          |           || | | | : | | : | :
Db      1119 CEDDVDECASQPCQHGGSCIDLVARYLCSCPPGTGLVLCIENEDDCGPGPPLDSGPRCLH 1178

Qy      504 GGACNTLDG--TCTCAPGWRGEKCEL---PCQDGTYGLNCAERCDCSHADGCHPTTG--- 555
          | | | | || | : | | : | | : | |
Db      1179 NGTCVDLVGGFRCTCPPGYTGLRCEADINECRSGA-----CHAAHTRDCLQDPGGGF 1230

Qy      556 HCRCLPGWSGVHCDSV-----CAEGRWGPNC 581
          | | | : | | : | | : | | |
Db      1231 RCLCHAGFSGPRCQTVLSPCESQPCQHGGQCRPSPGPGGGLTFTCHCAQPFWGPRC 1286

```

RESULT 7

A46019

notch-1 protein - mouse

N;Alternate names: motch protein

C;Species: Mus musculus (house mouse)

C;Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 07-Mar-2003

C;Accession: A46019; S25144; C49175; B46438; A46438; PH1569; S32109

R;del Amo, F.F.; Gendron-Maguire, M.; Swiatek, P.J.; Jenkins, N.A.; Copeland, N.G.; Gridley, T.

Genomics 15, 259-264, 1993

A;Title: Cloning, analysis, and chromosomal localization of Notch-1, a mouse homolog of Drosophila Notch.

A;Reference number: A46019; MUID:93194170; PMID:8449489

A;Accession: A46019

A;Status: not compared with conceptual translation

A;Molecule type: nucleic acid

A;Residues: 1-2531

A;Cross-references: GB:Z11886; GB:S47228; NID:g288502; PIDN:CAA77941.1; PID:g288503

A;Note: sequence extracted from NCBI backbone (NCBIP:127318)

R;Franco del Amo, F.; Smith, D.E.; Swiatek, P.J.; Gendron-Maguire, M.; Greenspan, R.J.; McMahon, A.P.; Gridley, T.

submitted to the EMBL Data Library, April 1992

A;Description: Expression pattern of Motch, a mouse homolog of Drosophila Notch, suggests an important role in early postimplantation mouse development.

A;Reference number: S25144

A;Accession: S25144

A;Molecule type: mRNA
 A;Residues: 1551-2108, 'Q', 2110-2114, 'ALP', 2118-2170 <FRA>
 A;Cross-references: EMBL:Z11886
 R;Lardelli, M.; Lendahl, U.
 Exp. Cell Res. 204, 364-372, 1993
 A;Title: Motch A and Motch B--two mouse Notch homologues coexpressed in a wide variety of tissues.
 A;Reference number: A49175; MUID:93178563; PMID:8440332
 A;Accession: C49175
 A;Status: preliminary; nucleic acid sequence not shown
 A;Molecule type: mRNA
 A;Residues: 1161-1547 <LAR>
 A;Cross-references: EMBL:X68278; NID:g287987; PIDN:CAA48339.1; PID:g287988
 A;Experimental source: embryo
 A;Note: sequence extracted from NCBI backbone (NCBIP:126159)
 R;Kopan, R.; Weintraub, H.
 J. Cell Biol. 121, 631-641, 1993
 A;Title: Mouse notch: expression in hair follicles correlates with cell fate determination.
 A;Reference number: A46438; MUID:93252998; PMID:8486742
 A;Accession: B46438
 A;Status: preliminary
 A;Molecule type: nucleic acid
 A;Residues: 1865-1932, 'RR', 1935-1937, 'L', 1938-1967, 'I', 1969-2044, 'IE', 2047-2052, 'S', 2054-2056, 'SIRRE', 2062-2075 <KOP>
 A;Experimental source: embryo
 A;Note: sequence extracted from NCBI backbone (NCBIN:131246, NCBIP:131247)
 C;Comment: This protein has many EGF repeats and lin-12[#1172]/Notch repeats.
 C;Comment: This protein is one of the neurogenic proteins controlling the decision between ectodermal and neural fate for cells in the early embryo.
 C;Genetics:
 A;Gene: notch-1
 A;Map position: 2
 A;Note: proximal region of chromosome 2
 C;Superfamily: notch protein; ankyrin repeat homology; EGF homology
 F;106-138/Domain: EGF homology <EGF1>
 F;144-175/Domain: EGF homology <EG01>
 F;222-254/Domain: EGF homology <EGF2>
 F;261-292/Domain: EGF homology <EG02>
 F;339-370/Domain: EGF homology <EG03>
 F;416-449/Domain: EGF homology <EGF3>
 F;456-487/Domain: EGF homology <EG04>
 F;494-525/Domain: EGF homology <EG05>
 F;532-563/Domain: EGF homology <EG06>
 F;607-638/Domain: EGF homology <EG07>
 F;682-713/Domain: EGF homology <EG08>
 F;757-788/Domain: EGF homology <EG09>
 F;795-826/Domain: EGF homology <EG10>
 F;873-904/Domain: EGF homology <EG11>
 F;911-942/Domain: EGF homology <EG12>
 F;949-980/Domain: EGF homology <EG13>
 F;987-1018/Domain: EGF homology <EG14>
 F;1025-1056/Domain: EGF homology <EG15>
 F;1063-1094/Domain: EGF homology <EG16>
 F;1149-1180/Domain: EGF homology <EG17>
 F;1187-1218/Domain: EGF homology <EG18>
 F;1233-1264/Domain: EGF homology <EGF4>

F;1352-1383/Domain: EGF homology <EG19>
 F;1391-1425/Domain: EGF homology <EGF>
 F;1917-1948/Domain: ankyrin repeat homology <AN1>
 F;1949-1981/Domain: ankyrin repeat homology <AN2>
 F;1983-2015/Domain: ankyrin repeat homology <AN3>
 F;2016-2048/Domain: ankyrin repeat homology <AN4>
 F;2049-2081/Domain: ankyrin repeat homology <AN5>

Query Match 19.2%; Score 693; DB 2; Length 2531;
 Best Local Similarity 25.3%; Pred. No. 4.7e-31;
 Matches 217; Conservative 71; Mismatches 206; Indels 364; Gaps 51;

```

Qy      10 SFICLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSCTDIILNWFKC 69
      |::|| | | :||:| | : | | :| | ::::|
Db      624 SYLCLCLKGTTGPNCIEINLDD---CA-----SNPCDS---GTCLDKIDGYEC 664

Qy      70 TRHRVSRYRTAYRHGEKTMYYRRKSQCCPGFYESGEMC---VPHCA----- 110
      | ||: :| || : ||
Db      665 A-----CEPGY--TGSMCNVNIDECAGSPCHNGGTCEDGIAG 699

Qy     111 -----DKCVHGRC---IAPNTCQCEPGWGGTNC---SSACDG 141
      : |||| | : | | ||| |||| :: |:
Db      700 FTCRCPEGYHDPTCLSEVNECNSNPCIHGACRDGLNGYKCDCAPGWSGTNCDINNNECES 759

Qy     142 DHWGPHTSRCQCKNGALCNPI TG--ACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCN 199
      : | | | | :| | | | | : | | :| | |
Db      760 N-----PCVNGGTCKDMTSGYVCTCREGFSGPNCQ-----TNINECASN-PCLN 802

Qy     200 GATC-DHVTG-ECRCPPGYTGAFCEDLCPGKGHPQCEQRCPCQNGGVC----HHVTGEC 253
      || | | | :| || |||| || : | | ||:| ||| : : |
Db      803 QGTCIDDVAGYKNCNCLPYTGATCEVVLAP-----C-ATSPCKNSGVCKESEDYESFSC 855

Qy     254 SCPSGWMGTVC-----GQPCPEGR-----FGKNCS---QECQ--- 282
      ||:|| | | || | | | | | :|| :|:
Db      856 VCPTGWQGQTCEVDINECVKSPCRHGASCQNTNGSYRCLCQAGYTGRNCESDIDDCRPNP 915

Qy     283 CHNGGTCDAA--TGQCHCSPGYTGERCQDE-----CPVGT 315
      |||||:| | | ||: | |::: ||||
Db      916 CHNGGSCTDGINTAFCDCPLPGFQGAFCEDINECASNPCQNGANCTDCVDSYTCTCPVGF 975

Qy     316 YGVLC AET-----CQCVNGGKCYHVS G----ACLCEAGFAGERCEARLCPEGLYGI-KC 364
      |: | | || | | | | | | | : | : :|
Db      976 NGIHCENNTPDCTESSCFNGGTC--VDGINSFTCLCPPGFTGSYCQ-----YDVNEC 1025

Qy     365 DKRCPCHLENTHSCHPM S G--ECACKPGWSGLYCNE-----TCSPGFYGEACQQI----- 412
      | | || | | :| | |::|| | :| | | |
Db     1026 DSR-PCLHGGT--CQDSYGYTKCTCPQGYTGLNCQNLVWRWCD SAPCKNGGRCWQTNTQYH 1082

Qy     413 CSCQN---GADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGIN-----CSSRCG----- 457
      | |:: | :|| :: | | :||: | | : : | : |
Db     1083 CECRSGWTGVNCDVLSVSCEVAAQKRIGDVTLLCQHGGLCVDEGDKHYCHCQAGYTGSYC 1142

Qy     458 -----CKNDAVCSPVDG--SCTCKAGWHGVDCS----- 483
      |:| | |: | || | ||:| | :||
Db     1143 EDEVDECSNPNCQNGATCTDYLGGFSCKCVAGYHGSNCSEEINECLSQPCQNGGT CIDLT 1202

Qy     484 ----IRCPSGTWGFGCNLT---C-----QCLNGGACNTLDG--TCTCAPGWRGE 523

```

```

      || || | | : | : || || | || || || : ||
Db      1203 NSYKCS CPRGTQGVHCEINVD DCHPPLDPASRSPKCFNNGTCVDQVGGYTCTCPPGFVGE 1262

Qy      524 KCE-----LPCQD-GTYGLNCAER-----CDC-----SHADGC----- 550
      :|| || || || :| :| | :||
Db      1263 RCEGDVNECLSNPCDPRGTQ--NCVQRVNDFHCECRA GHTGRRCESVINGCRGKPKCKNGG 1320

Qy      551 -----HPTTGH-CRCLPGWSGVHCDS-----VCAEGRWGPNC----- 581
      : | || | : | | : : | | | |
Db      1321 VCAVASNTARGFICRC PAGFEGATCENDARTCGSLRCLNGGTCISGPRSPTCLCLGSFTG 1380

Qy      582 -----SLPCY 586
      | |||
Db      1381 PECQFPASSPCVGSNPCY 1398

```

RESULT 8

A49175

Motch B protein - mouse (fragment)

N;Alternate names: Notch homolog

C;Species: Mus musculus (house mouse)

C;Date: 21-Jan-1994 #sequence_revision 05-Jan-1996 #text_change 08-Sep-2002

C;Accession: A49175; PH1570; S32113

R;Iardelli, M.; Lendahl, U.

Exp. Cell Res. 204, 364-372, 1993

A;Title: Motch A and Motch B--two mouse Notch homologues coexpressed in a wide variety of tissues.

A;Reference number: A49175; MUID:93178563; PMID:8440332

A;Accession: A49175

A;Status: preliminary; nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1-1203 <LAR>

A;Cross-references: EMBL:X68279; NID:g287989; PIDN:CAA48340.1; PID:g287990

A;Experimental source: embryo

A;Note: sequence extracted from NCBI backbone (NCBIP:126158)

C;Comment: This protein has many EGF repeats and lin-12/Notch repeats.

C;Comment: This protein is one of the neurogenic proteins controlling the decision between ectodermal and neural fate for cells in the early embryo.

C;Superfamily: notch protein; ankyrin repeat homology; EGF homology

F;143-174/Domain: EGF homology <EGX1>

F;482-513/Domain: EGF homology <EGF1>

F;560-591/Domain: EGF homology <EGF>

F;674-705/Domain: EGF homology <EGX2>

F;712-743/Domain: EGF homology <EGF3>

F;836-867/Domain: EGF homology <EGX3>

Query Match 19.1%; Score 687; DB 2; Length 1203;

Best Local Similarity 24.8%; Pred. No. 6.3e-31;

Matches 221; Conservative 78; Mismatches 245; Indels 348; Gaps 56;

```

Qy      3 ISLNSCLSFICL----LLCH-----WIGTASPLNLE--DPNVCSHW-----ES 39
      | : : | | | | : | : | | : | | : |
Db      214 IDIDDCSSTPCLNGAKCIDHPNGYECQCATGFTGILCDENIDNCDPDPCHHGQCQDGIDS 273

Qy      40 YSVTVQESYPHPF--DQI--YYTS-----CTDILNWFKCT-----RHRVSY----- 76
      | : | | | | | : | | : | | : : :
Db      274 YTCICNPGYMGAI CSDQIDECYSSPCLNDGRCIDL VNGYQCNCQPGTSGLNCEINFDDCA 333

```


Qy 77 RTAYRHG--EKTMYRRKSQCCPGF-----YESGEMCV----- 106
 || : | ||| | :
 Db 334 SNPCM HGVCVDGINRYSCVCS PGFTGQRCNIDIDE CASNPCRKGATCINDVNGFRICICPE 393

Qy 107 ----PHC-----ADKCVHGRC---IAPNTCQCEPGWGGTNCSSACDGDHWGPHCTSR 151
 | | : | : || | : | | | | | : | |
 Db 394 GPHHPSCYSQVNECLSNPCIHGNCTGGLSGYKCLCDAGWVGVNCE--VDKN----ECLSN 447

Qy 152 CQCKNGALCNPITGA--CHCAAGFRGWRCE---DRC-----EQGTYGNDCH-QRCQCQ- 198
 | : || | | : | | || : | : | | | | : | |
 Db 448 -PCQNGGTCNNLVNGYRCTCKKGFKGYNQVNIDECASNPCLNQGTCTFDDVSGYTCHCML 506

Qy 199 --NGATCDHVTGECRCPPGYTGAFCE-----LCPPGKHGPQCE---QRC---PCQ 241
 | | | | | | : | | | | | : | |
 Db 507 PYTGKNCQTVLAPCSPNPCENAAVCKEAPNFESFSLCAPGWQGKRCTVDVDECISKPCM 566

Qy 242 NGGVCHHVTGE--CSCPSGWMGTVCQPCPEGRFGKNCSQEC---QCHNGGTC--DAATG 294
 | ||| : | | | | : | | : | | | | : |
 Db 567 NNGVCHNTQGSYVCECPPGFGSGMDCEEDI-----NDCLANPCQNGGSCVDHVNTF 616

Qy 295 QCHCSPGYTGERCQDE-----CPVGTYGVLG---AETC-- 324
 | | || : | : || : | | : | | : |
 Db 617 SCQCHPGFIGDKCQTDMECLSEPCKNGGTCS DYVNSYTCTCPAGFHGVHCENNIDECTE 676

Qy 325 -QCVNGGKCYHVS-----ACLCEAGFAG-----ERCEAR-----LC 355
 | || | | | : || | | | : | : | : |
 Db 677 SSCFNNGTC--VDGINSFSLCPVGFTGPFCLHDINECSSNPCLNAGTCVDGLGTYRCIC 734

Qy 356 PEGLYGIKCD-----KRCPCHELENTHSCHPMSGECACKPGWSGLYCNE----- 398
 | | | | | | | : | | | | | | : |
 Db 735 PLGYTGKNCQTLVNLCSRSPCKNKGTQVQEKARPHCLCPPGWDGAYCDVLNVSKAAALQ 794

Qy 399 -----TCSPGFYGEACQ---ICS---CQNGADCDSVTG-- 426
 | | : | | : : | : | : || | : |
 Db 795 KGVPEHLHCQHSGICINAGNTHHCQCPGTYGTSYCEEQLDECASNPCQHGATCNDFIGGY 854

Qy 427 KCTCAPGFGKIDCSTPCPLGTYGINCSSRCGCKNDVAVCS PVDGSCCTCKAGWHGVDCSIRC 486
 : | | || : : || | : : | : | : | |
 Db 855 RCECVPGYQGVNCE-----YEVEDECQNQPCQNGGTCIDLNVNHFKCS-----C 896

Qy 487 PSGTWGFGC--NL-TC---QCLNGGAC--NTLDG--TCTCAPGWRGEKCE-----LPC 529
 | | | | | : | | | | | : : | | | | : || : ||
 Db 897 PPGTRGLLCEENIDECAGGPHCLNGGQCVDRIGGYTCRCLPGFAGERCEGDINECLSNPC 956

Qy 530 -QDGTYGLNCAE-----RCDCSHA-----DGC-----HPTTGHC 557
 : | : | : | | | | | | | | | |
 Db 957 SSEGS--LDCVQLKNNYNCICRSAFTGRHCETFLDVCPQKPCLNNGGTCAVASNMPDGFIC 1014

Qy 558 RCLPGWSGVHCDVCAEGRW-----GPNC-----SLPC 585
 || || : || | | : : | | | | | |
 Db 1015 RCPPGFSGARCQSSCGQVKRRGEQCIHTDSGPRCFCLNPKDCESGCASNPC 1066

RESULT 9
 S45306
 notch 3 protein - mouse

C;Species: Mus musculus (house mouse)
 C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 02-Aug-2002
 C;Accession: S45306
 R;Lardelli, M.; Dahlstrand, J.; Lendahl, U.
 Mech. Dev. 46, 123-136, 1994
 A;Title: The novel Notch homologue mouse Notch 3 lacks specific epidermal growth factor-repeats and is expressed in proliferating neuroepithelium.
 A;Reference number: S45306; MUID:95001556; PMID:7918097
 A;Accession: S45306
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-2318 <LAR>
 A;Cross-references: EMBL:X74760; NID:g483580; PIDN:CAA52776.1; PID:g483581
 C;Superfamily: notch protein; ankyrin repeat homology; EGF homology
 F;163-195/Domain: EGF homology <EGF1>
 F;474-505/Domain: EGF homology <EGF>
 F;854-885/Domain: EGF homology <EGF2>
 F;1839-1871/Domain: ankyrin repeat homology <AN1>
 F;1872-1904/Domain: ankyrin repeat homology <AN2>
 F;1906-1938/Domain: ankyrin repeat homology <AN3>
 F;1939-1971/Domain: ankyrin repeat homology <AN4>
 F;1972-2004/Domain: ankyrin repeat homology <AN5>

Query Match 19.0%; Score 685.5; DB 2; Length 2318;
 Best Local Similarity 24.4%; Pred. No. 1.2e-30;
 Matches 216; Conservative 59; Mismatches 195; Indels 417; Gaps 48;

```

Qy      94  CCPGFYESGEMCVPHCADKCVHGRCIAPNT-----CQCEPGWGGTNCSSACDGDHW 144
      | |||  | : | : | |  ||:  |  ||| | | | | :  |
Db      225 CLPGF--EGQNCEVN-VDDCPGHRCLNGGTCVDGVNTYNCQCPPEWTGQFCTEDVD---- 277

Qy      145 GPHCTSRCQ-----CKNGALCNPITG--ACHCAAG-----FRGWRCE 179
      ||  | | | | : | : | |  |  | | |
Db      278 -----ECQLQPNACHNGGTGCFNLLGGHSCVCVNGWTGESCSQNIDDCATAVCFHGATCH 331

Qy      180 DR-----CEQGTYGNDCH--QRC---QCQNGATCD--HVTGE--CRCPPGYTGAFCEDE 223
      ||  | | | | | |  |  |  | | | | : | | | | | :
Db      332 DRVASFYCACPMGKTGLLCHLDDACVSNPCHEDAICDTNPVSGRAICTCPPGFTGGACDQ 391

Qy      224 -----LCPPGKHGPQCE----- 235
      |  |  ||: ||
Db      392 DVDECSIGANPCEHLGRCVNTQGSFLCQCGRGYTGPRCETDVNECLSGPCRNQATCLDRI 451

Qy      236 -----QRCPCQNGGVC-HHVTG-ECSCPSGWMGTVC----- 264
      |  || ||||  |  |  | : ||| : | : |
Db      452 GQFTCICMAGFTGTYCEVDIDECQSSPCVNGGVCKDRVNGFSCCTCPSGFSGSMCQLDVDE 511

Qy      265 -----GQP-----CPEGRFGKNCSQ---ECQ---CHNGGTCDATGQCHC 298
      ||  | | | | :  : |  ||: |  | :  | |
Db      512 CASTPCRNGAKCVDQPDGYECRAEGFEGTLCERNVDDCS PDPCHHGRCVDGFIASFSCAC 571

Qy      299 SPGYTGERCQDE-----CPVGTYGVL-----AETCQ 325
      : |||| | : :  | | | | |  |  |
Db      572 APGYTGIRCESQVDECRSQPCRYGGKCLDLVDKYLRCPPGTTGVNCEVNIDDCASNPT 631

Qy      326 ---CVNGGKCYHVSAGLCEAGFAGERCEARL-----CPEGLYGIKCDKRCP 369
      | : |  |  | : : | | |  |  :  | : | |  | |
  
```

Db	632	FGVCRDGINRYD----CVCQPGFTGPLCNVEINECASSPCGEGGSCVDGENGFHC--LCP	685
Qy	370	-----CHLENTHS--CHPMMSG--ECACKPGWSGLYCNE-----	398
		: : :	
Db	686	PGSLPPLCLPANHPCAHKPCSHGVCHDAPGGFRCVCEPGWSGPRCSQSLAPDACESQPCQ	745
Qy	399	-----TCSPGFYGEACQQI--CS--CQNGADCDSVTGK--CTCAPGFKG	436
		: : : : : : :	
Db	746	AGGTCTSDGIGFRCTCAPGFQGHQCEVLSPECTPSLCEHGGHCESDPDRLTVCSCPPGWQG	805
Qy	437	-----IDCSTPCPLGTYGINCSS-----RCGCKNDAV-----CSP-----	466
		: : : :	
Db	806	PRCQQDVDEECAGASPCGPHG-TCTNLPGNFRICHRGYTGPFCDQDIDDCDPNPCLHGG	864
Qy	467	-VDG----SCTCKAGWHGVDC-----SIRCPSGTWGFGCNL	497
		: : :	
Db	865	CQDGVGSFSCSCLDGFAGPRCARDVDECLSSPCGPGTCTDHVASFTCACPPGYGGFHCEI	924
Qy	498	-----TCQCLNGGACNTLDG----TCTCAPGWRGEKC-----	525
		: :	
Db	925	DLPDCSPSSCFNGGTC--VDGVSSFSCLCRPGYTGTHCQYEADPCFSRPCLHGGICNPTH	982
Qy	526	---ELPCQDGTYGLNCAERCD-----CSHADGCHPTTGHCRCLPGWSGVHCD-----	569
		: : :	
Db	983	PGFECTCREGFTGSQCQNPVDWCSQAPCQNGGRCVQTGAYCICPPGWSGRLCDIQSLPCT	1042
Qy	570	-----SVCAEGRWGPNCSL--PC	585
		:	
Db	1043	EAAAQMGVRLEQLCQEGGKCIDKGRSHYCVCEPGRGTGSHCEHEVDPC	1089

RESULT 10

A49128

cell-fate determining gene Notch2 protein - rat

C;Species: Rattus norvegicus (Norway rat)

C>Date: 21-Jan-1994 #sequence revision 18-Nov-1994 #text_change 02-Aug-2002

C;Accession: A49128

R; Weinmaster, G.; Roberts, V.J.; Lemke, G.

Development 116, 931-941, 1992

A;Title: Notch2: a second mammalian Notch gene.

A:Reference number: A49128; MUID:93202015; PMID:1295745

A;Accession: A49128

A; Status: preliminary; not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-2471 <WEI>

A; Experimental source: Schwann cell

A;Note: sequence extracted from NCBI backbone (NCBIP:127811)

C;Superfamily: notch protein; ankyrin repeat homology; EGF homology

F:264-295/Domain: EGF homology <EGX1>

F;799-830/Domain: EGF homology <EGF1>

F:877-908/Domain: EGF homology <EGX2>

F:1029-1060/Domain: EGF homology <EGF>

F;1067-1098/Domain: EGF homology <EGX3>

F:1153-1184/Domain: EGF homology <EGF3>

F:1191-1222/Domain: EGF homology <EGX4>

F:1876-1908/Domain: ankyrin repeat homology <AN1>

F:1909-1941/Domain: ankyrin repeat homology <AN2>

F;1943-1975/Domain: ankyrin repeat homology <AN3>
 F;1976-2008/Domain: ankyrin repeat homology <AN4>
 F;2009-2041/Domain: ankyrin repeat homology <AN5>

Query Match 19.0%; Score 685.5; DB 2; Length 2471;
 Best Local Similarity 24.8%; Pred. No. 1.2e-30;
 Matches 218; Conservative 83; Mismatches 240; Indels 339; Gaps 55;

```

Qy      3 ISLNSCLSFICL----LLCH-----WIGTASPLNLE--DPNVCSHW-----ES 39
      | :: | | | : | : | | :: | | : |
Db      531 IDIDDCSSTPCLNGAKCIDHPNGYECQCATGFTGTLCDENIDNCDPDPCHHGQCQDGIDS 590

Qy      40 YSVTVQESYPHPF--DQI--YYTS-----CTDILNWFKCT-----RHRVSY---- 76
      | : | | | | : | | : | : | : :
Db      591 YTCICNPGYMGAIKSDQIDECYSSPCLNDGRCIDLVNGYQCNCQPGTSGLNCEINFDDCA 650

Qy      77 RTAYRHGE--KTMYYRRKSQCCPGFYESGEMC---VPHCADK----- 112
      | | : | | | : | : |
Db      651 SNPCLHGACVDGINRYSCVCSPGF--TGQRCNIDIDECASNPCRKDATCINDVNGFRMC 708

Qy     113 -----CVHGRC---IAPNTCQCEPGWGGTNCSSACDGDHWGPHCT 149
      | : | | : : | | : | | | | : |
Db      709 PEGPHHPSCYSQVNECLSSPCIHGNTGGLSGYKCLCDAGWVGINCE--VDKN----ECL 762

Qy     150 SRCQCKNGALCNPITGA--CHCAAGFRGWRCE---DRC-----EQGTYGNDCH-QRCQC 197
      | : | | | : | | | : | : | | | : |
Db      763 SN-PCQNGGTCNNLVNGYRCTCKKGFKGYNCQVNIDECASNPCLNQGTCLDDVSGYTCHC 821

Qy     198 Q---NGATCDHVTGECRCPPGYTGAFCE-----LCPPGKHGPQCE---QRC---P 239
      | | | | | | : | | | | : | | |
Db      822 MLPYTGKNCQTVLAPCSNPNCENAAVCKEAPNFESFTCLCAPGWQGGQRCTVDVDECVSKP 881

Qy     240 CQNGGVCHHVTGE--CSCPSGWMGTVCQGQPCPEGRFGKNCSQEC---QCHNGGTC--DAA 292
      | | : | : | | | : | | : : | | | : |
Db      882 CMNNGICHNTQGSYMCECPPGFSGMDCEEDI-----NDCLANPCQNGGSCVDKVN 931

Qy     293 TGQCHCSPGYTGERCQDE-----CPVGTYGVLG---AETC 324
      | | | | : | : | : | | : | | | : |
Db      932 TFSCLCLPGFVGDKCQTDMECLSEPCKNGGTCSDYVNSYTCTCPAGFHVHCENNIDEC 991

Qy     325 ---QCVNGGKCYHVS-----ACLCEAGFAGERC-----EARLCPEGLYGIKC 364
      | | | | | : | | | | : | : | : |
Db      992 TESSCFNGGTC--VDGINSFSLCPVGFTGPFCLHDINECSSNPCLNSGTCDVGLGTYRC 1049

Qy     365 -----DKRC-----PCHLENTHSCHPMSGECACKPGWSGLYCNE----- 398
      | | | : | : | | | | | :
Db     1050 TCPLGYTGKNCQTLVNLCSPPCKNKGTCAQEKARPRCLCPPGWDGAYCDVLNVSKAAA 1109

Qy     399 -----TCSPGFYGEACQQ---ICS---CQNGADCDSVTG 426
      | | : | : : | : | : | |
Db     1110 LQKGVPEVHLCQHSGICINAGNTHHCQCPGLGYTGSYCEEQLDECASNPCQHGAATCSDFIG 1169

Qy     427 --KCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSFVDGSCTCKAGWHGVDCSI 484
      : | | | : : : | : : | : |
Db     1170 GYRCECVPGYQGVNCE-----YEVDECQNQPCQNGGTCIDLNVNHFKCS----- 1212

Qy     485 RCPSGTWGFGC--NL-TC---QCLNGGAC-NTLDG-TCTCAPGWRGEKCE-----L 527

```

```

      || || | | | : | |||| | : : | : | ||: ||: ||
Db      1213 -CPPGTRGLLCEENIDDCAGAPHCLNGGQCVDRI GGYSRCRLPGFAGERCEGDINECLSN 1271

Qy      528 PC-QDGTYGLNCAE-----RCDCSHA-----DGCH-----PTTG 555
      || :|: |:| : :| | | | | | | |
Db      1272 PCSSEGS--LDCIQLKNNYQCVCRSAFTGRHCETFLDVCPQKPCLNGGTCAVASNVPDGF 1329

Qy      556 HCRCLPGWSGVHCD SVCAEGRW-----GPNCSLP 584
      ||| ||: || | | | : : ||: | |
Db      1330 ICRCPPGFSGARCQSSCGQVKCRRGEQCVHTASGPHCFCP 1369

```

RESULT 11

A40701

tenascin-X precursor - human

C;Species: Homo sapiens (man)

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Dec-1999

C;Accession: A40701; A33725; C42175

R;Bristow, J.; Tee, M.K.; Gitelman, S.E.; Mellon, S.H.; Miller, W.L.

J. Cell Biol. 122, 265-278, 1993

A;Title: Tenascin-X: a novel extracellular matrix protein encoded by the human

XB gene overlapping P450c21B.

A;Reference number: A40701; MUID:93300909; PMID:7686164

A;Accession: A40701

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-3566 <BRI>

A;Cross-references: EMBL:X71937

R;Morel, Y.; Bristow, J.; Gitelman, S.E.; Miller, W.L.

Proc. Natl. Acad. Sci. U.S.A. 86, 6582-6586, 1989

A;Title: Transcript encoded on the opposite strand of the human steroid 21-hydroxylase/complement component C4 gene locus.

A;Reference number: A33725; MUID:89367293; PMID:2475872

A;Accession: A33725

A;Molecule type: mRNA

A;Residues: 2748-3199,'V',3201-3298,'E',3299-3314,'G',3316-3566 <MOR>

A;Cross-references: GB:M25813; NID:g183069; PIDN:AAA35884.1; PID:g183070

R;Matsumoto, K.; Arai, M.; Ishihara, N.; Ando, A.; Inoko, H.; Ikemura, T.

Genomics 12, 485-491, 1992

A;Title: Cluster of fibronectin type III repeats found in the human major histocompatibility complex class III region shows the highest homology with the repeats in an extracellular matrix protein, tenascin.

A;Reference number: A42175; MUID:92217969; PMID:1373119

A;Accession: C42175

A;Molecule type: DNA

A;Residues: 1849-1936 <MAT>

A;Experimental source: clone 3.9kF3-1

A;Note: sequence extracted from NCBI backbone (NCBIP:95694)

C;Genetics:

A;Gene: GDB:TNXA; D6S103E; TNX; XA; XB

A;Cross-references: GDB:568487; OMIM:600261

A;Map position: 6p21.3-6p21.3

C;Superfamily: tenascin-X; EGF homology; fibrinogen beta/gamma homology; fibronectin type III repeat homology

C;Keywords: extracellular matrix; glycoprotein

F;435-461/Domain: EGF homology <EGF>

F;748-828/Domain: fibronectin type III repeat homology <3F1>

F;829-856/Domain: fibronectin type III repeat homology #status atypical <3F2>
 F;873-953/Domain: fibronectin type III repeat homology <3F3>
 F;975-1055/Domain: fibronectin type III repeat homology <3F4>
 F;1078-1158/Domain: fibronectin type III repeat homology <3F5>
 F;1167-1247/Domain: fibronectin type III repeat homology <3F6>
 F;1248-1317/Domain: fibronectin type III repeat homology #status atypical <3F7>
 F;1323-1403/Domain: fibronectin type III repeat homology <3F8>
 F;1412-1492/Domain: fibronectin type III repeat homology <3F9>
 F;1510-1590/Domain: fibronectin type III repeat homology <3F10>
 F;1618-1676/Domain: fibronectin type III repeat homology #status atypical <3F11>
 F;1678-1749/Domain: fibronectin type III repeat homology #status atypical <3F12>
 F;1751-1831/Domain: fibronectin type III repeat homology <3F13>
 F;1849-1929/Domain: fibronectin type III repeat homology <3F14>
 F;1955-2035/Domain: fibronectin type III repeat homology <3F15>
 F;2061-2141/Domain: fibronectin type III repeat homology <3F16>
 F;2167-2246/Domain: fibronectin type III repeat homology <3F17>
 F;2274-2354/Domain: fibronectin type III repeat homology <3F18>
 F;2382-2462/Domain: fibronectin type III repeat homology <3F19>
 F;2488-2568/Domain: fibronectin type III repeat homology <3F20>
 F;2584-2664/Domain: fibronectin type III repeat homology <3F21>
 F;2677-2757/Domain: fibronectin type III repeat homology <3F22>
 F;2771-2851/Domain: fibronectin type III repeat homology <3F23>
 F;2878-2958/Domain: fibronectin type III repeat homology <3F24>
 F;2977-3067/Domain: fibronectin type III repeat homology #status atypical <3F25>
 F;3078-3159/Domain: fibronectin type III repeat homology <3F26>
 F;3167-3247/Domain: fibronectin type III repeat homology <3F27>
 F;3255-3334/Domain: fibronectin type III repeat homology <3F28>
 F;3349-3557/Domain: fibrinogen beta/gamma homology <FBG>

Query Match 18.9%; Score 682; DB 1; Length 3566;
 Best Local Similarity 28.4%; Pred. No. 2.3e-30;
 Matches 191; Conservative 41; Mismatches 178; Indels 262; Gaps 38;

Qy	94	CCPGFYESG-----EMCVPHCADKCVHG-----RCIAPNTCQCEPGWGGTNCSSACDG	141
		: : :	
Db	125	CCPASAQAGTGQTDVRTLC-----SLHGVFDLSRC-----TCSCEPGWGGPTCSDPTDA	173
Qy	142	D-----HWGPHC---TSRCQCKNGALCNPITG	165
		: : : : :	
Db	174	EIPSSPPSASGSCPDDCNDQGRVCVRGRCVCFPGYTGPSCGWPSCPGDCQGRGRC--VQG	231
Qy	166	ACHCAAGFRGWRCEDR-CEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAF-ED	223
		: : :	
Db	232	VCVCRAFGSPDCSQRSCPRG-----CSQRGRCEG-----GRCVCDPGYTGDDCGMR	278
Qy	224	LCPPGKHGPQCEQRCPCQNGGVCHVTGECSCPSGWMGTVCG-QPCPEGRFGKNCSQECQ	282
		: : : :	
Db	279	SCPRG-----CSQRGRGEN-----GRCVCNPGYTGEDCGVRSCPRG-----CSQRGR	320
Qy	283	CHNGGTCDAAATGQCHCSPGYTGERC-QDECPVGTYGVLCAETCQCVNGGKCYHVSACLC	341
		: : : :	
Db	321	CKD-----GRCVCDPGYTGEDCGTRSCP-----WDCGEGGRC--VDGRCVC	359
Qy	342	EAGFAGERCEARLCPE-----GLYGIKCDKR-CPCHLENTHSCHPMSG	383
		:	
Db	360	WPGYTGEDCSTRTCPRDCRGRGRCEDEGEICDTGYSGDDCGVRSCPGDCNQGRGRC--DG	417

Qy 384 ECACKPGWSGLYCNE-----TCSPGFYGEAC-QQIC--SCQNGADCD 422
 | | ||:| | | : | | : | :
 Db 418 RVCWPGYTGTDCGSRACPRDCRGRGRCENGVCVCNAGYSGEDCGVRSCPGDCRGRGRCE 477

Qy 423 SVTGKCTCAPGFKGIDCST-----PCPLGTYGINCSS-RC--GCKND 461
 | | :| | ||: | || | | :| | || | :
 Db 478 S--GRCMCWPGYTGRDCGTRACPGDCRGRGRCDGRCVCNPGFTGEDCGSRRCPGDCRGRH 535

Qy 462 AVCSFVDGSCCTCKAGWHGVDCSIR-CPSGTWGFGCNLTCQCLNG----- 504
 :| | | | | | : | | | | | | | | | | :| :|
 Db 536 GLCE--DGVCVCDAGYSGEDCSTRSCPGGCRGRG-----QCLDGRVCVEDGYSGEDCGVR 588

Qy 505 -----GACNTLDGTCTCAPGWRGEKCELP-----CQDGTYGLN 537
 | | | | | | : | | : | | |
 Db 589 QCPNDCSQHGVQC--DGVCICWEGYVSEDCSIRTCPSNCHGRGRCEEGRCLCDPGYTGP 646

Qy 538 CAER---CDCSHADGCHPTTGHCRCLPGWSGVHC-----DSV 571
 || | | | | | : | | |
 Db 647 CATRMCPADCRGRGRG--VQGVCLCHVGYGGEDCGQEPPASACPGGCGPRELCRAGQCV 704

Qy 572 CAEGRWGPNC SL 583
 | | | ||:| :|
 Db 705 CVEGFRGPDCAI 716

RESULT 12

T31070

notch homolog - sea urchin (*Lytechinus variegatus*)

C;Species: *Lytechinus variegatus* (variegated urchin)

C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 31-Jan-2000

C;Accession: T31070

R;Sherwood, D.R.; McClay, D.R.

Development 124, 3363-3374, 1997

A;Title: Identification and localization of a sea urchin Notch homologue: insights into vegetal plate regionalization and Notch receptor regulation.

A;Reference number: Z20966; MUID:97454256; PMID:9310331

A;Accession: T31070

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-2531 <SHE>

A;Cross-references: EMBL:AF000634; NID:g2570350; PID:g2570351; PIDN:AAB82088.1

C;Superfamily: notch protein; ankyrin repeat homology; EGF homology

Query Match 18.9%; Score 681.5; DB 2; Length 2531;
 Best Local Similarity 24.2%; Pred. No. 2e-30;
 Matches 216; Conservative 68; Mismatches 215; Indels 393; Gaps 53;

Qy 6 NSCLSFICLLCHWIGTASPLNLED--PNVCSHWESYSVTVQESYPHPFDQIYYTSCTDI 63
 | : : | : : | : : | : | : |
 Db 336 NTYGNFSCICVRGWEGQTCEINKDDCTPNPCQ-----FEGECEDR 375

Qy 64 LNWFKCTRRHVSRYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRC----- 118
 : |||| | || :| :| | | :|
 Db 376 VAFKCT-----CPPG--RTGLLC--HLEDACMSNPCHHTAQ 408

Qy 119 ----IAPNT--CQCEPGWGGTNCSSACDGDHWGPHCTSRCQ--CKNGALCNPITG--ACH 168
 : : | | | : | || | | : | :| | | :|

Db 409 CSTSVVDGSFICDCATGYQGFCNSEDID-----ECSLSMDSICQSGGTCQNFDDGGWSCL 462
 Qy 169 CAAGFRGWRC---DRCE-----QGTYG-----NDCHQRCQ 196
 |::||| | || | | : : | | : |
 Db 463 CSSGFTGSRCETDIDECDDDPCYNGGTCLNKRGGYACICLTGFTGTLCETDINECSSN-P 521
 Qy 197 CQNGATCDHVTG--ECRCPPGYTGAFCE-----DLCPGPK 229
 | |||:| :|| || | |||| | : || |
 Db 522 CLNGASCFDITGRFECACLAGYTGTTCQVNIDDCQSSPCENGGTCIDGVNQFTCLCETGY 581
 Qy 230 HGPQCE---QRC---PCQNGGVCHHVTG--ECSCPSGWMGTVC-----GQP----- 267
 | :|| | || |||| : | :||| | | | |
 Db 582 EGHRCMDSDECASRPCMNGGVCEDLIGFYQCNCVPVGTSGDNCEYNHYDCSSNPCVNDGT 641
 Qy 268 -----CPEGRFGKNCSQ---ECQ---CHNGGTC-DAATG----- 294
 | || | || : :| : ||||| | | |
 Db 642 CVDGINEYTCMCHEGYRGLNCEEDIDDCESRPCHNGGTCVDEVNGYHCLCPIGYHDPFCM 701
 Qy 295 -----QCHCSPGYTGERCQ---DECPVGTYGVLCAETCQCQV 327
 | | |||| | || | : |
 Db 702 SNINECSSNPCVNGGSGCHDGVNEYSCECMAGYTGTCTDDDFDEC-----SSNPCQ 751
 Qy 328 NGGKC--YHVSAGACLCEAGFAGERCEARL-----CPEGLYG 361
 :|| | | | :||: | || : ||: |
 Db 752 HGGTCDNRHAFYNCTCQAGYTGLNCEVNIDDCVDEPCLNGGICIDEVNSFQCVCPQTFVG 811
 Qy 362 IKCD-KRCPCHLENTHSCHPMMSGACKPGWSGLYCNETCSPGFYGEACQ-----ICS 414
 : | : :| || : | | : :| | | : || | |
 Db 812 LLCETERSPC---EDNQCQ-NGATCVYSEDYAGYSCR--CTSGFQGNFCDDDRNECLFSP 865
 Qy 415 CQNGADCDSVTG--KCTCAPGFKGIDC-----STP-----CPLGT 447
 |::| | : : | :|| ||: | | | | |
 Db 866 CRNGGSCTNLEGSFECSCLPGYDGPICEINIDECASGPCTNGGICTDLIDDYFCSCQRGF 925
 Qy 448 YGINC---SSRC---GCKNDAVCSP-VDG-SCTCKAGWHGVDCSIRCPSGTWGFGCNLTC 499
 | || : | :| | | || :|| | : | : | | |
 Db 926 TGKNCQNDTDECLSSPCRNATCHEYVDSYTCSCIVGFSGMHCEINDQDCT-----TS 978
 Qy 500 QCLNGGACNTLDG----TCTCAPGWRGEKCEL-----PCQDGTYGLNCAER----- 541
 || || | :|| || | :| | : : ||:| | :|
 Db 979 SCLYGGTC--IDGVNSYTCECVTGYTGSNCQIEINECDSDPCENGA---TCQDREFGSYSC 1033
 Qy 542 -CD-----CSH-ADGCHP-----TTGH---CRCLPGWSGVHCD----- 569
 || | | | | || | | | | |
 Db 1034 HCDVGFGLNCEHVQWCSPQNNPCYNGATCVAMGHLYECHCASNWIGKLCDVPKVSCDI 1093
 Qy 570 -----SVCAEGRWGPNC-----SLPCY 586
 : | : | | | | : |
 Db 1094 AASDKNVTRSELCLNGGTCTIDATSSHSCLCQDGYTGSYCEVNIDECASAPCH 1145

RESULT 13

A24420

notch protein - fruit fly (*Drosophila melanogaster*)

N;Alternate names: neurogenic repetitive locus protein

C;Species: *Drosophila melanogaster*

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C;Accession: A24420; A24768; S09358; A05267
 R;Kidd, S.; Kelley, M.R.; Young, M.W.
 Mol. Cell. Biol. 6, 3094-3108, 1986
 A;Reference number: A24420; MUID:87064624; PMID:3097517
 A;Accession: A24420
 A;Molecule type: DNA
 A;Residues: 1-2703 <KID>
 A;Cross-references: GB:K03508; NID:g157991; PIDN:AAA28725.1; PID:g157993
 R;Wharton, K.A.; Johansen, K.M.; Xu, T.; Artavanis-Tsakonas, S.
 Cell 43, 567-581, 1985
 A;Reference number: A24768; MUID:86079539; PMID:3935325
 A;Accession: A24768
 A;Molecule type: mRNA
 A;Residues: 1-48,'I',50-118,'R',120-230,'I',232-256,'N',258-266,'A',268-
 872,'R',874-958,'R',960-1970,'FH',1973-2256,'G',2258-2264,'V',2266-
 2406,'R',2408-2444,'L',2446-2703 <WHA1>
 A;Note: the authors translated the codon ATC for residue 49 as Thr, ATT for
 residue 2044 as Arg, GTA for residue 2265 as Ala, CGC for residue 2407 as His,
 and CTT for residue 2445 as Arg
 R;Tautz, D.
 Nucleic Acids Res. 17, 6463-6471, 1989
 A;Title: Hypervariability of simple sequences as a general source for
 polymorphic DNA markers.
 A;Reference number: S09358; MUID:89385974; PMID:2780284
 A;Accession: S09358
 A;Molecule type: DNA
 A;Residues: 2505-2551,'QQQQ',2552-2576,'E',2578-2604 <TAU>
 R;Wharton, K.A.; Yedvobnick, B.; Finnerty, V.G.; Artavanis-Tsakonas, S.
 Cell 40, 55-62, 1985
 A;Title: opa: a novel family of transcribed repeats shared by the Notch locus
 and other developmentally regulated loci in D. melanogaster.
 A;Reference number: A05267; MUID:85099329; PMID:2981631
 A;Accession: A05267
 A;Molecule type: DNA
 A;Residues: 2504-2576,'E',2578-2611 <WHA2>
 C;Genetics:
 A;Gene: notch; opa
 A;Cross-references: FlyBase:FBgn0004647
 A;Map position: 8.96-9.36
 A;Introns: 53/3; 84/3; 171/3; 240/3; 283/3; 2333/3; 2436/3; 2588/3
 C;Superfamily: notch protein; ankyrin repeat homology; EGF homology
 C;Keywords: differentiation; tandem repeat; transmembrane protein
 F;27-43/Domain: transmembrane #status predicted <TMM1>
 F;297-328/Domain: EGF homology <EGX1>
 F;530-561/Domain: EGF homology <EGF1>
 F;568-599/Domain: EGF homology <EGF>
 F;988-1019/Domain: EGF homology <EGX2>
 F;1064-1095/Domain: EGF homology <EGF3>
 F;1187-1218/Domain: EGF homology <EGX3>
 F;1746-1762/Domain: transmembrane #status predicted <TMM2>
 F;1950-1982/Domain: ankyrin repeat homology <AN1>
 F;1983-2015/Domain: ankyrin repeat homology <AN2>
 F;1988-2004/Domain: transmembrane #status predicted <TMM3>
 F;2017-2049/Domain: ankyrin repeat homology <AN3>
 F;2050-2082/Domain: ankyrin repeat homology <AN4>
 F;2083-2115/Domain: ankyrin repeat homology <AN5>
 F;2538-2568/Region: glutamine-rich

F;2538-2568/Domain: neurogenic repetitive element #status predicted <OPA>

Query Match 18.8%; Score 677; DB 1; Length 2703;
Best Local Similarity 25.4%; Pred. No. 3.7e-30;
Matches 208; Conservative 78; Mismatches 203; Indels 330; Gaps 51;

```
Qy      7 SCL----SFICLLLLCHWIGTASPLNLED--PNVCSHWESYSVTVQESYPHPFDQIYYTSC 60
      |||      :| | : : | |      ::::| |      : :|
Db      502 SCLDDPGTFRVCVCMFGFTGTQCEIDIDECQSNPC-----LNDGTC 541

Qy      61 TDILNWFKCTRHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMC---VPHCADKCVHGR 117
      | :| |||:      | || :| | : | : |
Db      542 HDKINGFKCS-----CALGF--TGARCQINIDDCQSQPCRNR 576

Qy      118 -----CIAPNTCQCEPGWGGTNCS---SACDGDHWGPHCTSRCQCKNGALCNPITG-ACH 168
      || :||| ||: ||| | : || :      | | : : |
Db      577 GICHDSIAGYSCECPPGYTGTSCEININDCDN-----PCHRGKCIDDVNSFKCL 626

Qy      169 CAAGFRGW-----RCEDR-----CEQGTYG-----NDCHQRCQ 196
      | | : | :      | :||      | : || |      | :||
Db      627 CDPGYTGYYICQKQINECESNPCQFDGHCCQDRVGSYYCQCQAGTSGKNCEVNVNECHSN-P 685

Qy      197 CQNGATC-DHVTG-ECRCPPGYTGAFCEIDLCPGKHGPQCEQRCPCQNGGVC-HHVTG-E 252
      | |||| | : : :| ||:| ||      | : :| || | || | | :
Db      686 CNNGATCIDGINSYKQCQVPGFTGQHCE-----KNVDECIS-SPCANNGVCIDQVNGYK 738

Qy      253 CSCPSGWMGTVC-----GQP-----CPEGRFGKNCS---QECQ-- 282
      | || | : |      |      || | || |      ||
Db      739 CECPRGFYDAHCLSDVDECASNPCVNEGRCEDGINEFICHCPPGYTGKRCELDIDECSSN 798

Qy      283 -CHNGGTC-DAATG-QCHCSPGYTGERCQ---DECPVGTYGVLCAETCQCVNGGKCY-HV 335
      | :||| | | | ||||: :| :|      | | ||| | |
Db      799 PCQHGGTCYDKLNAFSCQCMPPGYTGQKCTETNIDDC-----VTNPCGNGGTCTIDKV 848

Qy      336 SG-ACLCEAGFAGERCEARLCPEGLYGIKC-DKRCPCHLENTHSCHPMSG----ECACKP 389
      :| | :| : | | ||: :| |      | || :| | | | | ||
Db      849 NGYKCVCKVPFTGRDCESKMDP-----CASNRC----KNEAKCTPSSNFLDFSCCKL 897

Qy      390 GWSGLYCNE-----TCSPGFYGEAC---QQICS---CQN 417
      | :| | ||:|      | :| | |      | : |||
Db      898 GYTGRYCDIDIDECSLSSPCRNASCLNVPGSYRCLCTKGYEGRDCAINTDDCASFPQN 957

Qy      418 GADCDSVTG--KCTCAPGFKGIDCST-----PCPLGTYGI 450
      | | | | | | | | | |      |||| ||
Db      958 GGTCLDGIGDYSCLCVDGFDGKHCETDINECLSQPCQNGATCSQYVNSYTCTCPLGFSGI 1017

Qy      451 NCS-----SRCGCKNDVAVCSPVDG----SCTCKAGWHGVDCSIR----- 485
      || : | | | :|| :|| ||: | :| :
Db      1018 NCQTNDEDCTESSCLNGGSC--IDGINGYNCSCLAGYSGANCQYKLNKCDNPNCLNGATC 1075

Qy      486 -----CPSGTWGFGCNL-----TCQCLNGGACNTL--DGTCTCAPGWRGEKCE- 526
      |||| | | :      | || | : : | | : || : |
Db      1076 HEQNNYETCHCPSGFTGKQCSEYVDWCGQSPCENGATCSQMKHQFSCKCSAGWTGKLCDV 1135

Qy      527 --LPCQDGT--YGLNCAERCD---CSHADGCHPTTGHCRCPLGWSGVHC----- 568
      : ||| || : : | | | | | | :| :|
Db      1136 QTISCQDAADRKGLSLRQLCNGGTCKDYGNSHV----CYCSQGYAGSYCQKEIDECQSQP 1191
```

Qy 569 -----DSVCAEGRWGPNCSL-----PC 585
: | : | | | | |
Db 1192 CQNGGTCDLIGAYECQCRQGFQGNCELNIDDCAPNPC 1230

RESULT 14

T09070

probable tenascin X - mouse

C;Species: Mus musculus (house mouse)

C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jan-2000

C;Accession: T09070

R;Rowen, L.; Mahairas, G.; Qin, S.; Ahearn, M.E.; Dankers, C.; Lasky, S.;
Loretz, C.; Schmidt, S.; Tipton, S.; Traicoff, R.; Zackrone, K.; Hood, L.
submitted to the EMBL Data Library, October 1997

A;Description: Sequence of the mouse major histocompatibility locus class III
region.

A;Reference number: Z16543

A;Accession: T09070

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-4006 <ROW>

A;Cross-references: EMBL:AF030001; NID:g2564945; PID:g2564958

C;Genetics:

A;Gene: TNX

A;Map position: 17

A;Introns: 124/1; 735/1; 773/3; 826/1; 914/1; 1037/1; 1135/1; 1232/1; 1329/1;
1440/1; 1548/1; 1645/1; 1745/1; 1841/1; 1938/1; 2045/1; 2154/1; 2253/1; 2367/1;
2476/1; 2589/1; 2696/1; 2804/1; 2911/1; 3019/1; 3115/1; 3208/1; 3302/1; 3405/1;
3517/1; 3558/1; 3606/1; 3646/1; 3694/1; 3737/3; 3782/1; 3832/3; 3865/1; 3919/1;
3973/3

C;Superfamily: tenascin-X; EGF homology; fibrinogen beta/gamma homology;
fibronectin type III repeat homology

C;Keywords: extracellular matrix

F;422-448/Domain: EGF homology <EGF>

F;826-906/Domain: fibronectin type III repeat homology <3FR>

F;3789-3997/Domain: fibrinogen beta/gamma homology <FBG>

Query Match 18.8%; Score 676.5; DB 2; Length 4006;
Best Local Similarity 29.1%; Pred. No. 5e-30;
Matches 190; Conservative 50; Mismatches 178; Indels 235; Gaps 38;

Qy 82 HGEKTMYYRRKSQCCPGF-----YESGEMCVPHCAD--KCVHGRCIAPNT 123
| | : | | | : : | | : | | | :
Db 135 HGVDLSRCACSCPEPGWGGPTCSDPTDTKTPTSSPPSKSCPEDCNDQGRVCVRGRCV---- 190

Qy 124 CQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRC 182
| | | : | : | | : | | | : | | | | | | |
Db 191 --CFPGYSGPSCSWPSCPGD-----CQGRGRC-----VKGVCVCRAFGFSG----PDC 231

Qy 183 EQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCE--EDLCPPGKHGFPQCEQRCPCQ 241
| : : | : | : | | | | | : | | | | | :
Db 232 SQRSRPRNCNQRCRCEE-----GRCVCDPGYSGEDCGVRSCPRG-----CSQRCRCE 278

Qy 242 NGGVCHHVTGECSCPSGWMGTVCG-QPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP 300
| | | | : | | | | | | | : | : | : | : | :
Db 279 N-----GLCVCNPGYSGEDCGVRNCPRG-----CSQRCRCED-----GRCVCDP 317

```

Qy      301 GYTGERCQDECPVGTYGVLCAETC--QCVNNGGKCYHVSGACLCEAGFAGERCEARLCPEG 358
      ||:|| |      || |:||:| | | |:| |:||| | | ||
Db      318 GYSGEDCS-----MRTCPWDCGDGGRC--VDGRCVCWPGYSGEDCSTRTCPRD 363

Qy      359 LYGI-KC-DKRCPCCHLE-----NTHSC-----HPMSGECACKPGWSGLYCNE---- 398
      | :| | | |      ||      | | | | |:| | :
Db      364 CRGRGRCEDGECICDAGYSGDDCGVRSCPGDCNQRGHCEDGRCVCWPGYTGADCSTRACP 423

Qy      399 -----TCSPGFYGEAC-QQIC--SCQNGADCDSVTGKCTCAPGFKGIDCST 441
      | |:|| | : | | : |:| | |:| | || |
Db      424 RDCRGRGRCEDGVCVCHAGYSGEDCGVRSCPGDCRGRNCES--GRCVCWPGYTGRDCGT 481

Qy      442 -----PCPLGTYGINCSS-RC--GCKNDAVCSPVDGSCCTCKAGWHGV 480
      | | | :| | || | : | | | | | : |
Db      482 RACPGDCRGRGRCDVGRVCNPGFTGEDCGSRRCPGDCRGHGHCE--NGVCVCAVGYSGD 539

Qy      481 DCSIR-CPSGTWGFGCNLTCQCLNG-----GACNTLDGTC 514
      ||| | ||| | | | |||| | | | || |
Db      540 DCSTRSCPSDCRGRG-----QCLNGLCECDEGYSGEDCGIRRCPRDCSQHGVCQ--DGLC 592

Qy      515 TCAPGWRGEKCEL-----PCQDGTYGLN-----CAER---CDCSHADGCHPT 553
      | | : || | : |:|| | | || | || | |
Db      593 MCHAGYAGEDCSIRTCPADCRRRGRCEDGRCVCNPGYTGPACATRTCPADCRGRGRG--V 650

Qy      554 TGHCRCLPGWSGVHC-----DSVCAEGRWGPNC SL 583
      | | | |:| | | | || | | |:| :
Db      651 QGVCMCYVGYSGEDCGQEPPASACPGGCGPRELCRAGQCVCVEGFRGPDCAI 703

```

RESULT 15

S18188

notch protein homolog - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 02-Aug-2002

C;Accession: S18188

R;Weinmaster, G.; Roberts, V.J.; Lemke, G.

Development 113, 199-205, 1991

A;Title: A homolog of Drosophila Notch expressed during mammalian development.

A;Reference number: S18188; MUID:92111383; PMID:1764995

A;Accession: S18188

A;Molecule type: mRNA

A;Residues: 1-2531 <WEI>

A;Cross-references: EMBL:X57405; NID:g57634; PID:g57635

C;Superfamily: notch protein; ankyrin repeat homology; EGF homology

F;987-1018/Domain: EGF homology <EGF1>

F;1025-1056/Domain: EGF homology <EGF>

F;1233-1264/Domain: EGF homology <EGF2>

F;1917-1949/Domain: ankyrin repeat homology <AN1>

F;1950-1982/Domain: ankyrin repeat homology <AN2>

F;1984-2016/Domain: ankyrin repeat homology <AN3>

F;2017-2049/Domain: ankyrin repeat homology <AN4>

F;2050-2082/Domain: ankyrin repeat homology <AN5>

Query Match 18.7%; Score 675; DB 2; Length 2531;

Best Local Similarity 25.7%; Pred. No. 4.6e-30;

Matches 208; Conservative 70; Mismatches 206; Indels 324; Gaps 48;

Qy 4 SLNSCLS-----FICLLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESY 48
 :| | | :| | | :| | | :| | |
 Db 603 NINECHSQPCRHHGGTCQDRDNYLCLCLKGTTPNCEINLDD---CA----- 646

Qy 49 PHFPDQIYYTSCTDILNWFKCTRHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMC--- 105
 :| | :| | :| | :| | :| | :| |
 Db 647 SNPCDS---GTCLDKIDGYECA-----CEPGY--TGSMCNVN 678

Qy 106 VPHCA-----DKCVHGRC---IAPNT 123
 : | | : | | | :
 Db 679 IDECAGSPCHNGGTCEGDIAGFTCRCPEGYHDPTCLSEVNECNSNPCIHGACRDGLNGYK 738

Qy 124 CQCEPGWGGTNC---SSACDGDHWGPHCTSRCQCKNGALCNPI TG--ACHCAAGFRGWRC 178
 | | | | | :| | :| | | | | | |
 Db 739 CDCAPGWSGTNCDINNECESN-----PCVNGGTCKDMTSGYVCTCREGFSGPNC 788

Qy 179 EDRCEQGTYGNDCHQRCQCGNGATC-DHVTG-ECRCPPGYTGAFCEDLCPGKHGPQCEQ 236
 : | | :| | | | | :| | | | | :| |
 Db 789 Q-----TNINECASN-PCLNQGTCTIDDVAGYKCNCLPYTGATCEVVLAP-----C-A 834

Qy 237 RCPCQNGGVC---HHVTGECSCPSGWMGTVC-----GQPCPEGR----- 272
 ||:| | | : : | | :| | | | | | |
 Db 835 TSPCKNSGVCKESEDYESFSCVCPTGWQGQTCEIDINECVKSPCRHGASCQNTNGSYRCL 894

Qy 273 -----FGKNCS---QECQ---CHNGGTCDAAATGQ--CHCSPGYTGERCQDE----- 310
 | :| | :| | | | | | | | | | :| | :| | :| |
 Db 895 CQAGYTGRNCESDIDDCRPNPCHNGGSGCTDGVNAAFCDCLPGFQGAFCCEEDINECATNPC 954

Qy 311 -----CPVGTYGVLCAET-----CQCVNGGKCYHVSG-----ACLCEAG 344
 | | | | :| | | | | | | | | | | |
 Db 955 QNGANCTDCVDSYTCTCPTGFNGIHCENNTPDCTESSCFNGGTC--VDGINSFTCLCPPG 1012

Qy 345 FAGERCEARLCPGLYGI-KCDKRCPCHLENTHSCHPMSG--ECACKPGWSGLYCNE--- 398
 | | | : | : :| | | | | | :| | | :| | | |
 Db 1013 FTGSYCQ-----YDVNECDNR-PC LHGGT--CQDSYGT YKCTCPQGYTGLNCQNLVR 1061

Qy 399 --TCSPGFYGEACQOI-----CSCQN---GADCDSVTGKCTCAPGFKGIDCSTPCPLGTY 448
 :| | | | | :| | :| | :| | :| | :| | :| |
 Db 1062 WCDSAPCKNGGKCWQTNTQYHCECRSGWTGFNCDVLSVSCEVAAQKRGIDVTLLCQHGGGL 1121

Qy 449 GIN-----CSSRCG-----CKNDAVCSPVDG--SCTCKAGWHGVDCS 483
 : : | : | :| | | :| | | | | :| | :| |
 Db 1122 CVDEEDKHYCHCQAGYTGSYCEDEVDECSNPNCQNGATCTDYLGGFSCKCVAGYHGSNCS 1181

Qy 484 -----IRCPSGTWGFGCNLT---C-----QCLNG 504
 | | | | | : | :| |
 Db 1182 EEINECLSQPCQNGGTCTDLTNTYKCS CPRGTQGVHCEINVDDCHPPLDPASRSPKCFNN 1241

Qy 505 GACNTLDG--TCTCAPGWRGEKCE-----LPCQD-GTYGLNCAERCDCSHADGCHPT 553
 | | | | | :| | :| | | | | :| | :| |
 Db 1242 GTCVDQVGGYTCTCPPGFVGERCEGDVNECLSNPCDPRGTQ--NCVQRVN----- 1289

Qy 554 TGHCRCLPGWSGVHCDVCAEGRWGPNC 581
 | | | | :| | :| | | | |
 Db 1290 DFHCECRAGHTGRRCESV-INGCRGKPC 1316

Search completed: March 26, 2004, 16:12:14

Job time : 19.883 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2004, 16:11:16 ; Search time 27.1611 Seconds
(without alignments)
5645.353 Million cell updates/sec

Title: US-10-092-390-4
Perfect score: 3601
Sequence: 1 MVISLNSCLSFICLLCHWI.....HCDSVCAEGRWGPNCSLPCY 586

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1065169 seqs, 261661801 residues

Total number of hits satisfying chosen parameters: 1065169

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query	Match Length DB ID	Description
No.				

1	3601	100.0	586	14	US-10-092-390-4	Sequence 4, Appli
2	3601	100.0	1140	14	US-10-092-390-2	Sequence 2, Appli
3	3601	100.0	1140	15	US-10-052-648A-33	Sequence 33, Appl
4	2094	58.2	969	15	US-10-052-648A-34	Sequence 34, Appl
5	2093	58.1	969	15	US-10-052-648A-35	Sequence 35, Appl
6	1851	51.4	1037	15	US-10-052-648A-10	Sequence 10, Appl
7	1830	50.8	1037	15	US-10-052-648A-8	Sequence 8, Appli
8	1828	50.8	1034	15	US-10-052-648A-31	Sequence 31, Appl
9	1824	50.7	1034	15	US-10-052-648A-32	Sequence 32, Appl
10	1811	50.3	1020	15	US-10-052-648A-2	Sequence 2, Appli
11	1799	50.0	928	15	US-10-052-648A-4	Sequence 4, Appli
12	1799	50.0	928	15	US-10-052-648A-6	Sequence 6, Appli
13	1667.5	46.3	1050	10	US-09-796-753-114	Sequence 114, App
14	1356.5	37.7	1393	12	US-10-312-352-21	Sequence 21, Appl
15	1284.5	35.7	1111	15	US-10-369-493-5280	Sequence 5280, Ap
16	1215	33.7	636	10	US-09-796-753-100	Sequence 100, App
17	1215	33.7	636	10	US-09-796-753-124	Sequence 124, App
18	1203	33.4	739	12	US-10-211-462-69	Sequence 69, Appl
19	1167	32.4	384	14	US-10-106-698-6253	Sequence 6253, Ap
20	1139.5	31.6	601	14	US-10-365-227-20	Sequence 20, Appl
21	1119.5	31.1	474	14	US-10-084-994-11	Sequence 11, Appl
22	1119.5	31.1	474	14	US-10-193-109-11	Sequence 11, Appl
23	1119.5	31.1	474	15	US-10-193-409-11	Sequence 11, Appl
24	843.5	23.4	830	9	US-09-870-759-134	Sequence 134, App
25	843.5	23.4	830	9	US-09-870-759-140	Sequence 140, App
26	843.5	23.4	830	10	US-09-842-758-57	Sequence 57, Appl
27	843.5	23.4	830	10	US-09-751-708A-134	Sequence 134, App
28	843.5	23.4	830	10	US-09-751-708A-140	Sequence 140, App
29	843.5	23.4	830	12	US-10-174-333-57	Sequence 57, Appl
30	808	22.4	865	10	US-09-842-758-20	Sequence 20, Appl
31	808	22.4	865	12	US-10-174-333-20	Sequence 20, Appl
32	808	22.4	866	12	US-10-433-579-4	Sequence 4, Appli
33	783.5	21.8	934	10	US-09-842-758-18	Sequence 18, Appl
34	783.5	21.8	934	12	US-10-174-333-18	Sequence 18, Appl
35	779	21.6	296	10	US-09-866-050A-458	Sequence 458, App
36	779	21.6	299	10	US-09-866-050A-192	Sequence 192, App
37	779	21.6	299	10	US-09-866-050A-332	Sequence 332, App
38	721.5	20.0	310	14	US-10-084-994-12	Sequence 12, Appl
39	721.5	20.0	310	14	US-10-193-109-12	Sequence 12, Appl
40	721.5	20.0	310	15	US-10-193-409-12	Sequence 12, Appl
41	719	20.0	2524	15	US-10-190-115-25	Sequence 25, Appl
42	719	20.0	2524	15	US-10-369-072-25	Sequence 25, Appl
43	708.5	19.7	2447	15	US-10-190-115-28	Sequence 28, Appl
44	708.5	19.7	2447	15	US-10-369-072-28	Sequence 28, Appl
45	697	19.4	2321	14	US-10-356-625-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
 US-10-092-390-4
 ; Sequence 4, Application US/10092390
 ; Publication No. US20030013865A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Yu, Xuanchuan

```
; APPLICANT: Miranda, Maricar
; TITLE OF INVENTION: No. US20030013865A1el Human EGF-Family Proteins and
Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0317-USA
; CURRENT APPLICATION NUMBER: US/10/092,390
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: US 60/275,013
; PRIOR FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 586
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-092-390-4
```

Qy	1	MVISLNSCLSFICLLLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC	60
Db	1	MVISLNSCLSFICLLLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC	60
Qy	61	TDILNWFKCTRHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIA	120
Db	61	TDILNWFKCTRHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIA	120
Qy	121	PNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCE	180
Db	121	PNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCE	180
Qy	181	RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPQCEQRCPC	240
Db	181	RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPQCEQRCPC	240
Qy	241	QNGGVCHHVTGECSCPSGWMGTVCQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP	300
Db	241	QNGGVCHHVTGECSCPSGWMGTVCQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP	300
Qy	301	GYTGERCQDECPVGTYGVLCAETCQCVNNGGKCYHVSAGACLCEAGFAGERCEARLCPEG	360
Db	301	GYTGERCQDECPVGTYGVLCAETCQCVNNGGKCYHVSAGACLCEAGFAGERCEARLCPEG	360
Qy	361	GIKCDKRCPCHELENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQIICSCQNG	420
Db	361	GIKCDKRCPCHELENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQIICSCQNG	420
Qy	421	CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCS PVDGSCTCKAGWHG	480
Db	421	CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCS PVDGSCTCKAGWHG	480
Qy	481	DCSIRCPSGTWGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPCQDGTYGLNCA	540
Db	481	DCSIRCPSGTWGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPCQDGTYGLNCA	540
Qy	541	RCDCSHADGCHPTTGHCRCPLGWSGVHCDSVCAEGRWGPNCSLPCY	586

Db 541 RCDCHADGCHPTTGHCRCLPGWSGVHCDSVCAEGRWGPNC SLPCY 586

RESULT 2

US-10-092-390-2

; Sequence 2, Application US/10092390

; Publication No. US20030013865A1

; GENERAL INFORMATION:

; APPLICANT: Yu, Xuanchuan

; APPLICANT: Miranda, Maricar

; TITLE OF INVENTION: No. US20030013865A1el Human EGF-Family Proteins and Polynucleotides Encoding the Same

; FILE REFERENCE: LEX-0317-USA

; CURRENT APPLICATION NUMBER: US/10/092,390

; CURRENT FILING DATE: 2002-03-06

; PRIOR APPLICATION NUMBER: US 60/275,013

; PRIOR FILING DATE: 2001-03-12

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 1140

; TYPE: PRT

; ORGANISM: homo sapiens

US-10-092-390-2

Query Match 100.0%; Score 3601; DB 14; Length 1140;
Best Local Similarity 100.0%; Pred. No. 4.4e-232;
Matches 586; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVISLNSCLSFICLLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC 60
|
Db 1 MVISLNSCLSFICLLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC 60

Qy 61 TDILNWFKCTRHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRCIA 120
|
Db 61 TDILNWFKCTRHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRCIA 120

Qy 121 PNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCED 180
|
Db 121 PNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCED 180

Qy 181 RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKGHPQCEQRCPC 240
|
Db 181 RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKGHPQCEQRCPC 240

Qy 241 QNGGVCHHVTGECSCPSGWMGTVCQGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP 300
|
Db 241 QNGGVCHHVTGECSCPSGWMGTVCQGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP 300

Qy 301 GYTGERCQDECPVGTYGVLCAETCQCVNNGGKCYHVS GACLCEAGFAGERCEARLCPEGLY 360
|
Db 301 GYTGERCQDECPVGTYGVLCAETCQCVNNGGKCYHVS GACLCEAGFAGERCEARLCPEGLY 360

Qy 361 GIKCDKRCPCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGAD 420
|
Db 361 GIKCDKRCPCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGAD 420

Qy	421	CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSVPDGSCTCKAGWHGV	480
Db	421	CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSVPDGSCTCKAGWHGV	480
Qy	481	DCSIRCPSGTWGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAE	540
Db	481	DCSIRCPSGTWGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAE	540
Qy	541	RCDCSHADGCHPTTGHCRCLPGWSGVHCDSVCAEGRWGPNCSLPCY	586
Db	541	RCDCSHADGCHPTTGHCRCLPGWSGVHCDSVCAEGRWGPNCSLPCY	586

RESULT 3

US-10-052-648A-33

; Sequence 33, Application US/10052648A

; Publication No. US20040005558A1

; GENERAL INFORMATION:

; APPLICANT: Anderson, David
 ; APPLICANT: Burgess, Catherine
 ; APPLICANT: Casman, Stacie
 ; APPLICANT: Colman, Steven
 ; APPLICANT: Edinger, Shlomit R.
 ; APPLICANT: Ellerman, Karen
 ; APPLICANT: Gerlach, Valerie
 ; APPLICANT: Gunther, Erik
 ; APPLICANT: Kekuda, Ramesh
 ; APPLICANT: MacDougall, John R.
 ; APPLICANT: Mehraban, Fuad
 ; APPLICANT: Patturajan, Meera
 ; APPLICANT: Rothenberg, Mark
 ; APPLICANT: Shimkets, Richard
 ; APPLICANT: Smithson, Glennnda
 ; APPLICANT: Spytek, Kimberly A.
 ; APPLICANT: Stone, David J.
 ; APPLICANT: Vernet, Corine A.M.
 ; APPLICANT: Zerhusen, Bryan D.

; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF

; TITLE OF INVENTION: USING THE SAME

; FILE REFERENCE: 21402-250 (CURA-550)

; CURRENT APPLICATION NUMBER: US/10/052,648A

; CURRENT FILING DATE: 2002-12-09

; PRIOR APPLICATION NUMBER: 60/262,454

; PRIOR FILING DATE: 2001-01-18

; PRIOR APPLICATION NUMBER: 60/272,920

; PRIOR FILING DATE: 2001-03-02

; PRIOR APPLICATION NUMBER: 60/284,549

; PRIOR FILING DATE: 2001-04-18

; PRIOR APPLICATION NUMBER: 60/303,229

; PRIOR FILING DATE: 2001-07-05

; PRIOR APPLICATION NUMBER: 60/262,892

; PRIOR FILING DATE: 2001-01-19

; PRIOR APPLICATION NUMBER: 60/263,605

; PRIOR FILING DATE: 2001-01-23

; PRIOR APPLICATION NUMBER: 60/269,098

; PRIOR FILING DATE: 2001-02-15

```

; PRIOR APPLICATION NUMBER: 60/264,159
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/271,855
; PRIOR FILING DATE: 2001-02-27
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 1140
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-052-648A-33

```

```

Query Match          100.0%; Score 3601; DB 15; Length 1140;
Best Local Similarity 100.0%; Pred. No. 4.4e-232;
Matches 586; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 MVISLNSCLSFICLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC 60
        ||||||||||||||||||
Db      1 MVISLNSCLSFICLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC 60

Qy     61 TDILNWFKCTRHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRCIA 120
        ||||||||||||||||||
Db     61 TDILNWFKCTRHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRCIA 120

Qy    121 PNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCED 180
        ||||||||||||||||||
Db    121 PNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCED 180

Qy    181 RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPFGKHGPQCEQRCPC 240
        ||||||||||||||||||
Db    181 RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPFGKHGPQCEQRCPC 240

Qy    241 QNGGVCHHVTGECSCPSGWMGTVCGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP 300
        ||||||||||||||||||
Db    241 QNGGVCHHVTGECSCPSGWMGTVCGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP 300

Qy    301 GYTGERCQDECPVGTYGVLCAETCQCVNNGKCYHVSGLCEAGFAGERCEARLCPEGLY 360
        ||||||||||||||||||
Db    301 GYTGERCQDECPVGTYGVLCAETCQCVNNGKCYHVSGLCEAGFAGERCEARLCPEGLY 360

Qy    361 GIKCDKRCPCHLNTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGAD 420
        ||||||||||||||||||
Db    361 GIKCDKRCPCHLNTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGAD 420

Qy    421 CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCS PVDGSC TCKAGWHGV 480
        ||||||||||||||||||
Db    421 CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCS PVDGSC TCKAGWHGV 480

Qy    481 DCSIRCPSGTWGFGCNLTQCCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAE 540
        ||||||||||||||||||
Db    481 DCSIRCPSGTWGFGCNLTQCCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAE 540

Qy    541 RCDCSHADGCHPTTGHCRCLPGWSGVHCDSVCAEGRWGPNCSLPCY 586
        ||||||||||||||||||

```

RESULT 4

US-10-052-648A-34

; Sequence 34, Application US/10052648A

; Publication No. US20040005558A1

; GENERAL INFORMATION:

; APPLICANT: Anderson, David

; APPLICANT: Burgess, Catherine

; APPLICANT: Casman, Stacie

; APPLICANT: Colman, Steven

; APPLICANT: Edinger, Shlomit R.

; APPLICANT: Ellerman, Karen

; APPLICANT: Gerlach, Valerie

; APPLICANT: Gunther, Erik

; APPLICANT: Kekuda, Ramesh

; APPLICANT: MacDougall, John R.

; APPLICANT: Mehraban, Fuad

; APPLICANT: Patturajan, Meera

; APPLICANT: Rothenberg, Mark

; APPLICANT: Shimkets, Richard

; APPLICANT: Smithson, Glennnda

; APPLICANT: Spytek, Kimberly A.

; APPLICANT: Stone, David J.

; APPLICANT: Vernet, Corine A.M.

; APPLICANT: Zerhusen, Bryan D.

; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF

; TITLE OF INVENTION: USING THE SAME

; FILE REFERENCE: 21402-250 (CURA-550)

; CURRENT APPLICATION NUMBER: US/10/052,648A

; CURRENT FILING DATE: 2002-12-09

; PRIOR APPLICATION NUMBER: 60/262,454

; PRIOR FILING DATE: 2001-01-18

; PRIOR APPLICATION NUMBER: 60/272,920

; PRIOR FILING DATE: 2001-03-02

; PRIOR APPLICATION NUMBER: 60/284,549

; PRIOR FILING DATE: 2001-04-18

; PRIOR APPLICATION NUMBER: 60/303,229

; PRIOR FILING DATE: 2001-07-05

; PRIOR APPLICATION NUMBER: 60/262,892

; PRIOR FILING DATE: 2001-01-19

; PRIOR APPLICATION NUMBER: 60/263,605

; PRIOR FILING DATE: 2001-01-23

; PRIOR APPLICATION NUMBER: 60/269,098

; PRIOR FILING DATE: 2001-02-15

; PRIOR APPLICATION NUMBER: 60/264,159

; PRIOR FILING DATE: 2001-01-25

; PRIOR APPLICATION NUMBER: 60/265,517

; PRIOR FILING DATE: 2001-01-31

; PRIOR APPLICATION NUMBER: 60/271,855

; PRIOR FILING DATE: 2001-02-27

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 97

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 34

; LENGTH: 969

; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-052-648A-34

Query Match 58.2%; Score 2094; DB 15; Length 969;
Best Local Similarity 65.4%; Pred. No. 1.2e-131;
Matches 312; Conservative 58; Mismatches 107; Indels 0; Gaps 0;

```
Qy      109 CADKCVHGRCIAPNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACH 168
      | :|||:|:|:|:| | | | | | :|| | | | | | :|:|:|:|:| | | | | |
Db      28 CTEECVHGRCVSPDTCHCEPGWGGPDCSSGCDSHDHWGPHCSNRCQCQNGALCNPITGACV 87

Qy      169 CAAGFRGWRCEDRCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCP PG 228
      | | | | | | | | : | | :| | | | | :|:|:|:| | | | | | :|:|:| | |
Db      88 CAAGFRGWRCCELCPGTHGKGCQLPCQCRHGASCDPRAGECLCAPGYTG VYCEELCP PG 147

Qy      229 KHGPQCEQRCPCQNGGVCHHVTGECSCPSGWMGTVCQPCPEGRFGKNCSQECQCHNGGT 288
      || | | | | | | | | :|:|:|:| | | | | | | | | | | :|:|:| | |
Db      148 SHGAHCELRCPCQNGGTCHHITGECACPPGWTGAVCAQPCPPGTFGQNCSDCPCHHGGQ 207

Qy      289 CDAATGQCHCSPGYTGERCQDECPVGTYGVLCAETCQCVNGGKCYHVS GACLCEAGFAGE 348
      || | | | | : | | | :|:|:| | | :|:| | | | | :|:| | | | |
Db      208 CDHVTGQCHCTAGYMGDRQCQECPFGSFGFQCSQRCDC HNGGQCSPTTGACECEPGYKGP 267

Qy      349 RCEARLCPEGLYGIKCDKRCPC HLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEA 408
      ||: | | | | :| | | | | :| | | | :| | | | :|:| | | | | :|:|:|
Db      268 RCQERLCPEGLHGP GCTLPCPCDADNTISCHPVTGACTCQPGWSGHHCNESCPVGYGDG 327

Qy      409 CQQICSCQNGADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSPVD 468
      || | :| | | | | | :| | | | | | | | : | | | | | | | | | | |
Db      328 CQLPCTCQNGADCHSITGGCTCAPGFMGEVCAVSCAAGTYGPNCSSICSCNNGGTCS PVD 387

Qy      469 GSCTCKAGWHGVDCSIRCPSGTWGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELP 528
      | | | | | | | | :|:|:| | | | | | | | :| | | | :|:|:| | | | |
Db      388 GSCTCKEGWQGLDCTLPCPSGTWGLNCNESCTCANGAACSPIDGSCSCTPGWLGDTC ELP 447

Qy      529 CQDGTYGLNCAERCDCSHADGCHPTTGHCRC LPGWSGVHCDSVCAEGRWGPNC SLPC 585
      | | | :|:|:|:| | | | | | | | | | | | | | | | | | | | | |
Db      448 CPDGTGFLNCSEHCDCSHADGCDPVTGHCCCLAGWTGIRCDSTCPPGRWGPNC SVSC 504
```

RESULT 5

US-10-052-648A-35

; Sequence 35, Application US/10052648A

; Publication No. US20040005558A1

; GENERAL INFORMATION:

; APPLICANT: Anderson, David
; APPLICANT: Burgess, Catherine
; APPLICANT: Casman, Stacie
; APPLICANT: Colman, Steven
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Ellerman, Karen
; APPLICANT: Gerlach, Valerie
; APPLICANT: Gunther, Erik
; APPLICANT: Kekuda, Ramesh
; APPLICANT: MacDougall, John R.
; APPLICANT: Mehraban, Fuad

```

; APPLICANT: Patturajan, Meera
; APPLICANT: Rothenberg, Mark
; APPLICANT: Shimkets, Richard
; APPLICANT: Smithson, Glennnda
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Stone, David J.
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Zerhusen, Bryan D.
; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF
; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 21402-250 (CURA-550)
; CURRENT APPLICATION NUMBER: US/10/052,648A
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/262,454
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/272,920
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: 60/284,549
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/303,229
; PRIOR FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 60/262,892
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/263,605
; PRIOR FILING DATE: 2001-01-23
; PRIOR APPLICATION NUMBER: 60/269,098
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: 60/264,159
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/271,855
; PRIOR FILING DATE: 2001-02-27
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 969
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (848)..(889)
; OTHER INFORMATION: Where Xaa is any amino acid
US-10-052-648A-35

```

```

Query Match          58.1%; Score 2093; DB 15; Length 969;
Best Local Similarity 65.4%; Pred. No. 1.4e-131;
Matches 312; Conservative 58; Mismatches 107; Indels 0; Gaps 0;

```

```

QY      109 CADKCVHGRCIAPNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACH 168
        | ::|||||:::| | ||||| :| | | |||||::|||:|||||
Db      28 CTEECVHGRCVSPDTCHEPGWGGPDCSSGCDSDHWGPHCSNRCQCQNGALCNPITGACV 87

QY      169 CAAGFRGWRCEDRCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPPG 228
        ||||| |||||: | ||: | |||||::| | ||| | |||: ||: ||||
Db      88 CAAGFRGWRCCELCPGTHGKGCQLPCQCRHGASCDPRAGECLCAPGYTGVYCEELCPPG 147

```

Qy	229	KHGPQCEQRCPCQNGGVCHHVTGECSCPSGWMGTVCQPCPEGRFGKNCSQECQCHNGGT	288
		: : : : :	
Db	148	SHGAHCELRPCPCQNGGTCHHITGECACPPGWTGAVCAQPCPPGTFGQNCSDCPCHHGGQ	207
Qy	289	CDAATGQCHCSPGYTGERCQDECPVGTYGVLCAETCQCVNNGKCYHVSGACLCEAGFAGE	348
		: : : : : : : : :	
Db	208	CDHVTGQCHCTAGYMGDRQCQEECPFGSFGFQCSQHCDCHNGGQCSPTTGACECEPGYKGP	267
Qy	349	RCEARLCPEGLYGIKCDKRCPCCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEA	408
		: : : : : : : : :	
Db	268	RCQERLCPEGLHGPGLTLPCCDADNTISCHPVTGACTCQPGWSGHHCNESCPVGYGDDG	327
Qy	409	CQQICSCQNGADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSPVD	468
		: : :	
Db	328	CQLPCTCQNGADCHSITGGCTCAPGFMGEVCAVSCAAGTYGPNCSSICSCNNGGTCSVD	387
Qy	469	GSCTCKAGWHGVDCSIRCPSTWGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELP	528
		: : : : : : : :	
Db	388	GSCTCKEGWQGLDCTLPCTSGTWGLNCNESCTCANGAACSPIDGSCSCTPGWLGDTCCLP	447
Qy	529	CQDGTGYLNCARCDCSHADGCHPTTGHCRCLPGWSGVHCDVCAEGRWGPNCVSLPC	585
		: : : : :	
Db	448	CPDGTGFLNCSEHCDCSHADGCDPVTGHCCCLAGWTGIRCDSTCPPGRWGPNCVSC	504

RESULT 6

US-10-052-648A-10

; Sequence 10, Application US/10052648A

; Publication No. US20040005558A1

; GENERAL INFORMATION:

; APPLICANT: Anderson, David

; APPLICANT: Burgess, Catherine

; APPLICANT: Casman, Stacie

; APPLICANT: Colman, Steven

; APPLICANT: Edinger, Shlomit R.

; APPLICANT: Ellerman, Karen

; APPLICANT: Gerlach, Valerie

; APPLICANT: Gunther, Erik

; APPLICANT: Kekuda, Ramesh

; APPLICANT: MacDougall, John R.

; APPLICANT: Mehraban, Fuad

; APPLICANT: Patturajan, Meera

; APPLICANT: Rothenberg, Mark

; APPLICANT: Shimkets, Richard

; APPLICANT: Smithson, Glennnda

; APPLICANT: Spytek, Kimberly A.

; APPLICANT: Stone, David J.

; APPLICANT: Vernet, Corine A.M.

; APPLICANT: Zerhusen, Bryan D.

; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF

; TITLE OF INVENTION: USING THE SAME

; FILE REFERENCE: 21402-250 (CURA-550)

; CURRENT APPLICATION NUMBER: US/10/052,648A

; CURRENT FILING DATE: 2002-12-09

; PRIOR APPLICATION NUMBER: 60/262,454

; PRIOR FILING DATE: 2001-01-18

```

; PRIOR APPLICATION NUMBER: 60/272,920
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: 60/284,549
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/303,229
; PRIOR FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 60/262,892
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/263,605
; PRIOR FILING DATE: 2001-01-23
; PRIOR APPLICATION NUMBER: 60/269,098
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: 60/264,159
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/271,855
; PRIOR FILING DATE: 2001-02-27
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
;   LENGTH: 1037
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-052-648A-10

```

```

Query Match          51.4%;   Score 1851;   DB 15;   Length 1037;
Best Local Similarity 52.0%;   Pred. No. 2e-115;
Matches 299;   Conservative 61;   Mismatches 209;   Indels 6;   Gaps 3;

```

```

Qy      14 LLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSCTDILNW---FKCT 70
      ||| : | || ||| || |||: | :||: || : | | |
Db      9 LLLAVGLRLAGTLNPSDPNTCSFEWESFTTTTKESHSRPFSLLPSEPCE--RPWEGPHTCP 66

Qy      71 RHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGW 130
      : | ||| || || :||: || ||||| | ||| || :||| ||| ||| |||
Db      67 QPTVVYRTVYRQVVKTDHRQLQCCHGFYESREFCVPLCAQECVHGRCAVAPNQCQCVPGW 126

Qy     131 GGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGND 190
      | :||| | ||| | | | | : | : | : | | : | : | | |
Db     127 RGDDCSSECAPGMWGPQCDKPCSCGNNSSCDPKSGVCSCPSGLQPPNCLQPCTPGYYGPA 186

Qy     191 CHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPQCEQRCPCQNGGVCHHVT 250
      | |||| :|| || || | || || | : | | | | | |||||
Db     187 CQFRCQC-HGAPCDPQTGACFCPAERTGPSCDVSCSQTSGFFCPSTHPCQNGGVFQTPQ 245

Qy     251 GECSCPSGWMGTVCQGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSPGYTGERCQDE 310
      | |||| |||||: | |||| | |||||: |||| | ||| | :||| | :||: |
Db     246 GSCSCPPGWMGTICSLPCEGFHGPNCSEQECRCHNGGLCDRFTGQCRCAPGYTGDRCREE 305

Qy     311 CPVGTYGVLCAETCQCVNGGKCYHVSAGACLCEAGFAGERCEARLCPEGLYGIKDKRCPC 370
      |||| :| |||| | :| : |||| | | :|| | ||| : | | | |
Db     306 CPVGRFGQDCAETCDAPDARCFPANGACLCEHGFTGDRCTDRLCPDGFYGLSCQAPCTC 365

Qy     371 HLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCDSVTGKCTC 430
      | : | ||||: |||: | |||: | |||: | |||: | | : | : | |

```


Db 366 DREHSLSCHPMNGECSCPLPGWAGLHCNESCPOQDTHGPGCQEYCLCLHGGVCQATSGLCQC 425
 Qy 431 APGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSPVDGSCCTCKAGWHGVDCSIRCPST 490
 |||: | |:| || |||: |||: || |:| |||: || | | | | :||: || ||
 Db 426 APGYTGPHCASLCPDPTYGVNCSARCSCENAIACSPIDGECVCKEGWQRGNCSVPCPPGT 485
 Qy 491 WGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPCQDGTYGLNCAERCDCHADGC 550
 ||| || :||| : |: | ||| ||| | |:||| | :| || |||| |:|||
 Db 486 WGFSCNASCQCAHEAVCSPQTGACTCTPGWHGAHCQLPCPKQFGEGCASRCDCHSDGC 545
 Qy 551 HPTTGHCRLPGWSGVHCDVCAEGRWGPNCSLPC 585
 | | |:| || | | | || || ||| |
 Db 546 DPVHGRCQCQAGWMGARCHLSCPEGLWGVNCSNTC 580

RESULT 7

US-10-052-648A-8

; Sequence 8, Application US/10052648A

; Publication No. US20040005558A1

; GENERAL INFORMATION:

; APPLICANT: Anderson, David
 ; APPLICANT: Burgess, Catherine
 ; APPLICANT: Casman, Stacie
 ; APPLICANT: Colman, Steven
 ; APPLICANT: Edinger, Shlomit R.
 ; APPLICANT: Ellerman, Karen
 ; APPLICANT: Gerlach, Valerie
 ; APPLICANT: Gunther, Erik
 ; APPLICANT: Kekuda, Ramesh
 ; APPLICANT: MacDougall, John R.
 ; APPLICANT: Mehraban, Fuad
 ; APPLICANT: Patturajan, Meera
 ; APPLICANT: Rothenberg, Mark
 ; APPLICANT: Shimkets, Richard
 ; APPLICANT: Smithson, Glennnda
 ; APPLICANT: Spytek, Kimberly A.
 ; APPLICANT: Stone, David J.
 ; APPLICANT: Vernet, Corine A.M.
 ; APPLICANT: Zerhusen, Bryan D.

; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF

; TITLE OF INVENTION: USING THE SAME

; FILE REFERENCE: 21402-250 (CURA-550)

; CURRENT APPLICATION NUMBER: US/10/052,648A

; CURRENT FILING DATE: 2002-12-09

; PRIOR APPLICATION NUMBER: 60/262,454

; PRIOR FILING DATE: 2001-01-18

; PRIOR APPLICATION NUMBER: 60/272,920

; PRIOR FILING DATE: 2001-03-02

; PRIOR APPLICATION NUMBER: 60/284,549

; PRIOR FILING DATE: 2001-04-18

; PRIOR APPLICATION NUMBER: 60/303,229

; PRIOR FILING DATE: 2001-07-05

; PRIOR APPLICATION NUMBER: 60/262,892

; PRIOR FILING DATE: 2001-01-19

; PRIOR APPLICATION NUMBER: 60/263,605

; PRIOR FILING DATE: 2001-01-23

; PRIOR APPLICATION NUMBER: 60/269,098

```
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: 60/264,159
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/271,855
; PRIOR FILING DATE: 2001-02-27
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 1037
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-052-648A-8
```

```
Query Match          50.8%; Score 1830; DB 15; Length 1037;
Best Local Similarity 51.7%; Pred. No. 5e-114;
Matches 297; Conservative 61; Mismatches 211; Indels 6; Gaps 3;
```

```
Qy      14 LLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSTCDILNW---FKCT 70
      | | | : | | | | | | | | | | | | | | | | | | | | | | |
Db      9 LLLAVGLRLAGTILNPSDENTCSEWESEFTTTKESHRSRPFSLLPSEPCE--RPWEGPHTCP 66

Qy      71 RHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNT'CQCEPGW 130
      : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      67 QPTVVYRTVYRQVVKTDHRQRLQCCGHGFYESRGFCVPLCAQECVHGRCVAPNQCQCVPGW 126

Qy     131 GGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGND 190
      | : | | | | | | | | | | | | | | | | | | | | | | | |
Db     127 RGDDCSSECAPGMWGPQCDKPCSCGNNSSCDPKSGVCSCPSGLQPPNCLQPCPTPGYYGPA 186

Qy     191 CHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPQCEQRCPCQNGGVCHHVT 250
      | | | | : | | | | | | | | | | | | | | | | | | | |
Db     187 CQFRCQC-HGAPCDPQTGACFCPAERTGPSCDVSCSQGTSGFFCPSTHPCQNGGVFQTPQ 245

Qy     251 GECSCPSGWMGTVCQGQPCPEGFRGKNCSEQECQCHNGGTCDAAATGQCHCSPGYTGERCQDE 310
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     246 GSCSCPPGWMGTICSLPCPEGFHGPNCSQECRCHNGGLCDRFTGQCRCAPGYTGDRCREE 305

Qy     311 CPVGTYGVLCATCQCQVNGGKCYHVSAGACLCEAGFAGERCEARLCPEGLYGIKCDKRCPC 370
      | | | | : | | | | | | | | | | | | | | | | | | | |
Db     306 CPVGRFGQDCAETCDCAPDARCFPANGACLCEHGFTGDRCTDRLCPDGFYGLSCQAPRTC 365

Qy     371 HLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCDSVTGKCTC 430
      | : | | | | | | | | | | | | | | | | | | | | | | | |
Db     366 DREHSLSCHPMNGECSCLPGWAGLHCNESCPQDTHGPGCQEHCLLHGGVCQATSGLCQC 425

Qy     431 APGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCS PVDGSC TCKAGWHGVDCSIRCPSGT 490
      | | | : | | | | | | | | | | | | | | | | | | | | |
Db     426 APGYTGPHCASLCPDPTYGVNCSARCSCENAIACSPIDGECVCKEGWQRGNCSVPCCPGT 485

Qy     491 WGFGCNLTQCCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAERCDCSHADGC 550
      | | | | : | | | : | | | | | | | | | | | | | | | |
Db     486 WGFSCNASCQCAHEAVCSPQTGACTCTPGWHGAHCQLPCPKGQFGECCASRCDCHSDGC 545

Qy     551 HPTTGHCRCCLPGWSGVHCDSVCAEGRWGPNC SLPC 585
```

Db 546 DPVHGRCQCQAGWMGARCHLSCPEGLWGVNCSNTC 580

RESULT 8

US-10-052-648A-31

; Sequence 31, Application US/10052648A

; Publication No. US20040005558A1

; GENERAL INFORMATION:

; APPLICANT: Anderson, David

; APPLICANT: Burgess, Catherine

; APPLICANT: Casman, Stacie

; APPLICANT: Colman, Steven

; APPLICANT: Edinger, Shlomit R.

; APPLICANT: Ellerman, Karen

; APPLICANT: Gerlach, Valerie

; APPLICANT: Gunther, Erik

; APPLICANT: Kekuda, Ramesh

; APPLICANT: MacDougall, John R.

; APPLICANT: Mehraban, Fuad

; APPLICANT: Patturajan, Meera

; APPLICANT: Rothenberg, Mark

; APPLICANT: Shimkets, Richard

; APPLICANT: Smithson, Glennnda

; APPLICANT: Spytek, Kimberly A.

; APPLICANT: Stone, David J.

; APPLICANT: Vernet, Corine A.M.

; APPLICANT: Zerhusen, Bryan D.

; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF

; TITLE OF INVENTION: USING THE SAME

; FILE REFERENCE: 21402-250 (CURA-550)

; CURRENT APPLICATION NUMBER: US/10/052,648A

; CURRENT FILING DATE: 2002-12-09

; PRIOR APPLICATION NUMBER: 60/262,454

; PRIOR FILING DATE: 2001-01-18

; PRIOR APPLICATION NUMBER: 60/272,920

; PRIOR FILING DATE: 2001-03-02

; PRIOR APPLICATION NUMBER: 60/284,549

; PRIOR FILING DATE: 2001-04-18

; PRIOR APPLICATION NUMBER: 60/303,229

; PRIOR FILING DATE: 2001-07-05

; PRIOR APPLICATION NUMBER: 60/262,892

; PRIOR FILING DATE: 2001-01-19

; PRIOR APPLICATION NUMBER: 60/263,605

; PRIOR FILING DATE: 2001-01-23

; PRIOR APPLICATION NUMBER: 60/269,098

; PRIOR FILING DATE: 2001-02-15

; PRIOR APPLICATION NUMBER: 60/264,159

; PRIOR FILING DATE: 2001-01-25

; PRIOR APPLICATION NUMBER: 60/265,517

; PRIOR FILING DATE: 2001-01-31

; PRIOR APPLICATION NUMBER: 60/271,855

; PRIOR FILING DATE: 2001-02-27

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 97

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 31

; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-052-648A-31

Query Match 50.8%; Score 1828; DB 15; Length 1034;
Best Local Similarity 52.0%; Pred. No. 6.8e-114;
Matches 299; Conservative 59; Mismatches 211; Indels 6; Gaps 3;

```
Qy      14 LLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSCTDILNW---FKCT 70
      ||| : || ||||: |||: | :||: || : || |
Db      7 LLLALGLRLTGTLNSNDPNVCTFWESFTTTTKESHLRPFSLLPAESCH--RPWEDPH'TCA 64

Qy      71 RHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGW 130
      : | ||| || | : ||| |:||| ||| || :|||:| || ||| |||
Db      65 QPTVVYRTVYRQVVKMDSRPRQLCCRGYYESRGACVPLCAQECVHGRCVAPNQCCAPGW 124

Qy     131 GGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGND 190
      | :||| | ||| | | | : | :| :||| | :| : | | |||
Db     125 RGGDCSSECAPGMWGPQCDKFCHCGNNSCDPKSGACFCPSGLQPPNCLQPCPAGHYGPA 184

Qy     191 CHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKGHPQCEQRCPCQNGGVCHHVT 250
      | ||| ||: || | |||| | | | | | | | | | | | | | | |
Db     185 CQFDCQCY-GASCDPQDGACFCPPGRAGPSCNVPCSQGTGDGFFCPRTYPCQNGGVPQGSQ 243

Qy     251 GECSCPSGWMGTVCQGQPCPEGRFGKNCSEQECQCHNGGTCDAATGQCHCSPGYTGERCQDE 310
      | |||| |||| :| |||| | ||: ||: |||| || |||||: ||| | :||: |
Db     244 GSCSCPPGWMGVICSLPCPEGFHGPNCQTQECRCHNGGLCDRFTGQCHCAPGYIGDRCQEE 303

Qy     311 CPVGTYGVLC AETCQCVNGGKCYHVS GACLCEAGFAGERCEARLCPEGLYGIKCDKRCPC 370
      |||| :| |||| | | :| : :||| || | :|| ||||: | ||: | :| |
Db     304 CPVGRFGQDCAETCDCAFGARCFPANGACLCEHGFTGDRCTERLCPDGRYGLSCQEPCTC 363

Qy     371 HLENTHSCHPMMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCDSVTGKCTC 430
      | :| |||| |||: |||: |||: |||: | :| ||: | | :| | : :| | |
Db     364 DPEHSLSCHPMHGECSQPGWAGLHCNESCPQDTHGPGCQEHCLCLHGGLCLADSGLCRC 423

Qy     431 APGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCS PVDGSCCTCKAGWHGVDCSIRCPSGT 490
      |||: | | : || ||||| ||| :| |||: ||: | || || :||: || ||
Db     424 APGYTGPHCANLCPDPTYGINCSSRCSCENAIACSPIDGTICKEGWQRGNCSPVCPPLGT 483

Qy     491 WGFGCNLTQCCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAERCDCSHADGC 550
      ||| || :||| : | | : | ||| ||| | | :||| | :| || ||| | :|||
Db     484 WGFNCNASCQCAHDGVCSPQTGACTCTPGWHGAHCQLPCPKGQFGEGCASVCDCHSDGC 543

Qy     551 HPTTGHCRCLPGWSGVHCDSVCAEGRWGPNC SLPC 585
      | | ||| || | | | || || || || |
Db     544 DPVHGQCRCQAGWMGTRCHLPCPEGEFGANCSNTC 578
```

RESULT 9
US-10-052-648A-32
; Sequence 32, Application US/10052648A
; Publication No. US20040005558A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; APPLICANT: Burgess, Catherine

```

; APPLICANT: Casman, Stacie
; APPLICANT: Colman, Steven
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Ellerman, Karen
; APPLICANT: Gerlach, Valerie
; APPLICANT: Gunther, Erik
; APPLICANT: Kekuda, Ramesh
; APPLICANT: MacDougall, John R.
; APPLICANT: Mehraban, Fuad
; APPLICANT: Patturajan, Meera
; APPLICANT: Rothenberg, Mark
; APPLICANT: Shimkets, Richard
; APPLICANT: Smithson, Glennnda
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Stone, David J.
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Zerhusen, Bryan D.
; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF
; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 21402-250 (CURA-550)
; CURRENT APPLICATION NUMBER: US/10/052,648A
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/262,454
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/272,920
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: 60/284,549
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/303,229
; PRIOR FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 60/262,892
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/263,605
; PRIOR FILING DATE: 2001-01-23
; PRIOR APPLICATION NUMBER: 60/269,098
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: 60/264,159
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/271,855
; PRIOR FILING DATE: 2001-02-27
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
;   LENGTH: 1034
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-10-052-648A-32

```

```

Query Match          50.7%;   Score 1824;   DB 15;   Length 1034;
Best Local Similarity 51.8%;   Pred. No. 1.3e-113;
Matches 298;   Conservative 59;   Mismatches 212;   Indels 6;   Gaps 3;

```

```

Qy      14 LLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSCTDILNW---FKCT 70
      |||      :      ||  |||||: |||:: | :||: ||  :   ||      |

```

Db 7 LLLALGLRLTGTLNSNDPNVCTFWESFTTTTKESHLRPFSLIPAESCH--RPWEDPHTCA 64

Qy 71 RHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGW 130
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 65 QPTVVYRTVYRQVVKMDSRPRLQCCRGYYESRGACVPLCAQECVHGRCVAPNQCCAPGW 124

Qy 131 GGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGND 190
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 125 RGGDCSSECAPGMWGPQCDKFCHCGNNSCDPKSGTCFCPSGLQPPNCLQPCPAGHYGPA 184

Qy 191 CHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPFGKHGPQCEQRCPCQNGGVCHHVT 250
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 185 CQFDCQCY-GASCDPQDGACFCPPGRAGPSCNVPCSQGTDFGFCPRTYPCQNGGVPQGSQ 243

Qy 251 GECSCPSGWMGTVCQGQPCPEGRFGKNCSEQCQCHNGGTCDAAATGQCHCSPGYTGERCQDE 310
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 244 GSCSCPPGWMGVICSLPCPEGFHGPNTQECRCHNGGLCDRFTGQCHCAPGYIGDRCQEE 303

Qy 311 CPVGTYGVLCAETCQCVNGGKCYHVSAGLCEAGFAGERCEARLCPEGLYGIKCDKRCPC 370
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 304 CPVGRFGQDCAETCDCAPGARCFPANGACLCEHGFTGDRCTERLCPDGRYGLSCQEPCTC 363

Qy 371 HLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCDSVTGKCTC 430
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 364 DPEHSLSCHPMHGECSQPGWAGLHCNESCQPDTHGPGCQEHCLCLHGGGLCLADSGLCRC 423

Qy 431 APGFGKIDCSTPCPLGTYGINCSSRCGCKNDVCSVPDVGSCCTKAGWHGVDCSIRCPSTG 490
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 424 APGYTGPHCANLCPDPTYGINCSSRCSCENAIACSPIDGTICKEGWQRGNCSVPCPLGT 483

Qy 491 WGFGCNLTQCCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAERCDCSHADGC 550
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 484 WGFENCNASCQCAHDGVCSPQTGACTCTPGWHGAHCQLPCPKGQFGEGCASVCDCHSDGC 543

Qy 551 HPTTGHCRCPLPGWSGVHCDSVCAEGRWGPNCSLPC 585
: | | | | | : | | | | | : | | | | | : | | | | |

Db 544 DPVHGQCRCQAGWMGTRCHLPCPEGFEGWGANCSNTC 578

RESULT 10

US-10-052-648A-2

; Sequence 2, Application US/10052648A

; Publication No. US20040005558A1

; GENERAL INFORMATION:

; APPLICANT: Anderson, David
; APPLICANT: Burgess, Catherine
; APPLICANT: Casman, Stacie
; APPLICANT: Colman, Steven
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Ellerman, Karen
; APPLICANT: Gerlach, Valerie
; APPLICANT: Gunther, Erik
; APPLICANT: Kekuda, Ramesh
; APPLICANT: MacDougall, John R.
; APPLICANT: Mehraban, Fuad
; APPLICANT: Patturajan, Meera
; APPLICANT: Rothenberg, Mark

```

; APPLICANT: Shimkets, Richard
; APPLICANT: Smithson, Glennnda
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Stone, David J.
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Zerhusen, Bryan D.
; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF
; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 21402-250 (CURA-550)
; CURRENT APPLICATION NUMBER: US/10/052,648A
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/262,454
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/272,920
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: 60/284,549
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/303,229
; PRIOR FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 60/262,892
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/263,605
; PRIOR FILING DATE: 2001-01-23
; PRIOR APPLICATION NUMBER: 60/269,098
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: 60/264,159
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/271,855
; PRIOR FILING DATE: 2001-02-27
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1020
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-052-648A-2

```

```

Query Match          50.3%; Score 1811; DB 15; Length 1020;
Best Local Similarity 49.4%; Pred. No. 9.1e-113;
Matches 297; Conservative 61; Mismatches 211; Indels 32; Gaps 4;

```

```

Qy      14 LLLCHWIGTASPLNLEDPNVC SHWESYSVTVQESYPHPFDQIYYTSCTDILNW---FKCT 70
      ||| : | || || |||:: | :||: || : | | |
Db      9 LLLAVGLRLAGTLNPSDPNTCSFWESFTTTTKESH SRPFSLLPSEPCE--RPWEGPHTCP 66

Qy      71 RHRVSYRTAYRHGEKTM YRRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGW 130
      : | ||| || || :||: || |||| | || || :|||::||| ||| |||
Db      67 QPTVVYRTVYRQVVKTDHRQLQCCHGFYESRGFCVPLCAQECVHGRCVAPNQCCVPGW 126

Qy     131 GGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGND 190
      | :||| | ||| | | | :|:| :| | :| : | | |
Db     127 RGDDCSSECAPGMWGPQCDKPKSCGNNSSCDPKSGVCSCPSGLQPPNCLQPCTPGYYGPA 186

Qy     191 CHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCP PGKHGPQCEQRCPCQNGGVCHHVT 250

```

Db	187	CQFRCQC-HGAPCDPQTGACFCPAERTGPSCDVSCSQGTSGFFCPSTHSCQNGGVFQTPQ	245
Qy	251	GECSGPSGWM-----GTVCGQPCPEGRFGKNCSQECQCH	284
Db	246	GSCSCPPGWMVWRVGFVGMGCGSGENSVGGAKQGSKGITCSLPCPEGFHHGNCSQECRCH	305
Qy	285	NGGTCDAAATGQCHCSPGYTGERCQDECVPVGTGVLCETCQCVNNGGKCYHVSAGACLCEAG	344
Db	306	NGGLCDRFTGQCRCAPGYTGDRCREECPVGRFGQDCAETCDCAPDARCFPANGACLCEHG	365
Qy	345	FAGERCEARLCPEGLYGIKCDKRCPCHELENTHSCHPMMSGECACKPGWSGLYCNETCSPGF	404
Db	366	FTGDRCTDRLCPDGFYGLSCQAPCTCDREHSLSCHPMNGECSCLPGWAGLHCNESCPQDT	425
Qy	405	YGEACQQICSCQNGADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDVAVC	464
Db	426	HGPCQEHCLCLHGGVCQATSGLCQCAPGYTGPHCASLCPDPTGVNCSARCSCEAIAIC	485
Qy	465	SPVDGSCTCKAGWHGVDCSIRCPSTWGFSGCNLTQCCLNGGACNTLDGTCTCAPGWRGEK	524
Db	486	SPIDGEVCVKEGWQRGNCSVPCPPGTWGFSCNASQCAHEAVCSPTGTACTCTPGWHGAH	545
Qy	525	CELPCQDGTGTYGLNCAERCDCSHADGCHPTTGHCRCLPGWSGVHCDVCAEGRWGPNCCLP	584
Db	546	CQLPCPKGQFGEGCASRCDCHSDGCDPVHGRQCQAGWMGARCHLSCEGLWGVNCSNT	605
Qy	585	C 585	
Db	606	C 606	

RESULT 11

US-10-052-648A-4

; Sequence 4, Application US/10052648A

; Publication No. US20040005558A1

; GENERAL INFORMATION:

; APPLICANT: Anderson, David
 ; APPLICANT: Burgess, Catherine
 ; APPLICANT: Casman, Stacie
 ; APPLICANT: Colman, Steven
 ; APPLICANT: Edinger, Shlomit R.
 ; APPLICANT: Ellerman, Karen
 ; APPLICANT: Gerlach, Valerie
 ; APPLICANT: Gunther, Erik
 ; APPLICANT: Kekuda, Ramesh
 ; APPLICANT: MacDougall, John R.
 ; APPLICANT: Mehraban, Fuad
 ; APPLICANT: Patturajan, Meera
 ; APPLICANT: Rothenberg, Mark
 ; APPLICANT: Shimkets, Richard
 ; APPLICANT: Smithson, Glennnda
 ; APPLICANT: Spytek, Kimberly A.
 ; APPLICANT: Stone, David J.
 ; APPLICANT: Vernet, Corine A.M.
 ; APPLICANT: Zerhusen, Bryan D.

; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF


```

; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 21402-250 (CURA-550)
; CURRENT APPLICATION NUMBER: US/10/052,648A
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/262,454
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/272,920
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: 60/284,549
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/303,229
; PRIOR FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 60/262,892
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/263,605
; PRIOR FILING DATE: 2001-01-23
; PRIOR APPLICATION NUMBER: 60/269,098
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: 60/264,159
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/271,855
; PRIOR FILING DATE: 2001-02-27
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 928
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-052-648A-4

```

```

Query Match          50.0%; Score 1799; DB 15; Length 928;
Best Local Similarity 49.3%; Pred. No. 5.3e-112;
Matches 296; Conservative 61; Mismatches 212; Indels 32; Gaps 4;

```

```

Qy      14 LLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSCDILNW---FKCT 70
      ||| : | || ||| || |||:: | :||: || : | | |
Db      9 LLLAVGLRLAGTLNPSDPNTCSFWESFTTTTKESHRSRPFSLLPSEPCE--RPWEGPHTCP 66

Qy      71 RHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGW 130
      : | ||| || || :|:: ||| ||||| ||| || :|||::||| ||| |||
Db      67 QPTVVYRTVYRQVVKTDHRQRLQCCHGFYESRGFCVPLCAQECVHGRCVAPNQCQCVPGW 126

Qy      131 GGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGND 190
      | :||| | ||| | ||| :|:| :| | | :| : | | |||
Db      127 RGDDCSSECAPGMWGPQCDKPCSCGNNSSCDPKSGVCSCPSGLQPPNCLQPCTPGYYGPA 186

Qy      191 CHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGHPQCEQRCPCQNGGVCHHVT 250
      | |||| :|| || || || || || : | | | | |||||
Db      187 CQFRCQC-HGAPCDPQTGACFCPAERTGPSCDVSCSQGTSGFFCPSTHSCQNGGVFQTPQ 245

Qy      251 GECSCPSGWM-----GTVCGQPCPEGRFGKNCSQECQCH 284
      | |||| ||| |||:| |||| | |||||:|
Db      246 GSCSCPPGWMVWRVGPVGMGCGSGENSVGGAKQSGKGTICSLPCEGFHGPNCSEQECRCH 305

```

Qy 285 NGGTCDAATGQCHCSPGYTGERCQDECPVGTYGVLCAETCQCVNNGGKCYHVS GACLCEAG 344
 ||| || |||| |:||||:|:|:|||| | |||| | :|: :||||| |
 Db 306 NGGLCDRFTGQCRCAPGYTGDRCREECFVGRFGQDCAETCDCAPDARCFPANGACLCEHG 365

Qy 345 FAGERCEARLCPEGLYGIKCDKRCPCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGF 404
 | |:|| ||||:| ||: | | |: ||||:|:|:| |||:|:|:|:|
 Db 366 FTGDRCTDRLCPDGFYGLSCQAPRTCDREHSLSCHPMNGECSCLPGWAGLHCNESCPODT 425

Qy 405 YGEACQQICSCQNGADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVC 464
 :| ||: | | :| | : :| | ||||: | |: || |||:|:|:| | |:| |
 Db 426 HGPQCQEHCLCLHGGVCQATSGLCQCAPGYTGPHCASLCPDPTYGVNCSARCSCENAIAC 485

Qy 465 SPVDGSCTCKAGWHGVDCSIRCPSGTWGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEK 524
 ||:|| | || || :||: || |||| || :||| : | : | ||| ||| |
 Db 486 SPIDGECVCKEGWQRGNCSPCPPGTWGFSCNASCQCAHEAVCSPQTGACTCTPGWHGAH 545

Qy 525 CELPCQDGTYGLNCAERCDCSHADGCHPTTGHCRCPLGWSGVHCDSVCAEGRWGPNCSLP 584
 |:||| | :| || |||| |:||| | | |:| || | | || ||| |||
 Db 546 CQLPCPKGQFGEGCASRCDCHSDGCDPVHGRCCQAGWMGARCHLSCPEGLWGVNCSNT 605

Qy 585 C 585
 |
 Db 606 C 606

RESULT 12

US-10-052-648A-6

; Sequence 6, Application US/10052648A

; Publication No. US20040005558A1

; GENERAL INFORMATION:

; APPLICANT: Anderson, David

; APPLICANT: Burgess, Catherine

; APPLICANT: Casman, Stacie

; APPLICANT: Colman, Steven

; APPLICANT: Edinger, Shlomit R.

; APPLICANT: Ellerman, Karen

; APPLICANT: Gerlach, Valerie

; APPLICANT: Gunther, Erik

; APPLICANT: Kekuda, Ramesh

; APPLICANT: MacDougall, John R.

; APPLICANT: Mehraban, Fuad

; APPLICANT: Patturajan, Meera

; APPLICANT: Rothenberg, Mark

; APPLICANT: Shimkets, Richard

; APPLICANT: Smithson, Glennnda

; APPLICANT: Spytek, Kimberly A.

; APPLICANT: Stone, David J.

; APPLICANT: Vernet, Corine A.M.

; APPLICANT: Zerhusen, Bryan D.

; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF

; TITLE OF INVENTION: USING THE SAME

; FILE REFERENCE: 21402-250 (CURA-550)

; CURRENT APPLICATION NUMBER: US/10/052,648A

; CURRENT FILING DATE: 2002-12-09

; PRIOR APPLICATION NUMBER: 60/262,454

; PRIOR FILING DATE: 2001-01-18

; PRIOR APPLICATION NUMBER: 60/272,920

```

; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: 60/284,549
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/303,229
; PRIOR FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 60/262,892
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/263,605
; PRIOR FILING DATE: 2001-01-23
; PRIOR APPLICATION NUMBER: 60/269,098
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: 60/264,159
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/271,855
; PRIOR FILING DATE: 2001-02-27
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 928
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-052-648A-6

```

```

Query Match          50.0%;   Score 1799;   DB 15;   Length 928;
Best Local Similarity 49.3%;   Pred. No. 5.3e-112;
Matches 296;   Conservative 61;   Mismatches 212;   Indels 32;   Gaps 4;

```

```

Qy      14 LLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSCTDILNW---FKCT 70
      ||| : | || ||| |||:: | :||: || : | | |
Db      9 LLLAVGLRLAGTLNPSDPNTCSFWESFTTTTKESHSRPFSLLPSEPCE--RPWEGPHTCP 66

Qy      71 RHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGW 130
      : | ||| || |||:|:: ||| |||| ||| |||:|||||:|||| ||| |||
Db      67 QPTVVYRTVYRQVVKTDHRQLQCCHGFYESRGFCVPLCAQECVHGRCVAPNQCCQCVPGW 126

Qy     131 GGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGND 190
      | :||| | ||| | | | | : | : | : | : | : | | | |
Db     127 RGDDCSSECAPGMWGPQCDKPCSCGNSSCDPKSGVCSCPSGLQPPNCLQPCTPGYYGPA 186

Qy     191 CHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPFGKHGPQCEQRCPCQNGGVCHHVT 250
      | |||| :|| || || | || || | : | | | | | |||||
Db     187 CQFRCQC-HGAPCDPQTGACFCPAERTGPSCDVSCSQGTSGFFCPSTHSCQNGGVFQTPQ 245

Qy     251 GECSCPSGWM-----GTVCGQPCPEGRFGKNCSQECQCH 284
      | |||| ||| ||:| |||| | |||||:|
Db     246 GSCSCPPGWMVWRVGPVGMGCGSGENSVGGAKQGSKGITCSLPCPEGFHGPNCSSQECRCH 305

Qy     285 NGGTCDAAATGQCHCSPGYTGERCQDECPVGTYGVLCAETCQCVNNGGKCYHVSACLCCEAG 344
      ||| || |||| |:|||||:|::| |||| :| |||| | :| : ||||| |
Db     306 NGGLCDRFTGQRCAPGYTGDRCREECPVGRFGQDCAETCDCAPDARCFPANGACLCCEHG 365

Qy     345 FAGERCEARLCPEGLYGIKCDKRCPCHELENTHSCHPMMSGECACKPGWSGLYCNETCSPGF 404
      | :|| ||||:| ||: | | | :| ||||:||||:| |||:|:|:|
Db     366 FTGDRCTDRLCPDGFYGLSCQAPRTCDREHSLSCHPMNGECSCLPGWAGLHCNESCPCQDT 425

```

Qy 405 YGEACQQICSCQNGADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVC 464
 :| ||: | | :| | : :| | ||||: | |:: || |||:||||:| | :| |
 Db 426 HGPGCQERCLCLHGGVCQATSGLCQCAPGYTGPHCASLCPDITYGVNCSARCSCEAIAIC 485
 Qy 465 SPVDGSCTCKAGWHGVDCSIRCPSGTWGFGCNLTQCQLNGGACNTLDGTCTCAPGWRGEK 524
 ||:|| | || | :||: || |||| | | :|| : | : | || | || |
 Db 486 SPIDGECVCKEGWQRGNCVPCPPGTWGFSCNASCQCAHEAVCSPQTGACTCTPGWHGAH 545
 Qy 525 CELPCQDGTYGLNCAERCDCSHADGCHPTTGHCRCPLPGWSGVHCDVCAEGRWGPNCSLP 584
 |:|| | :| || |||| |:||| | | |:| || | | | || || ||
 Db 546 CQLPCPKGQFGEGCASRCDCHSDGCDPVHGRCQCQAGWMGARCHLSCPEGLWGVNCSNT 605
 Qy 585 C 585
 |
 Db 606 C 606

RESULT 13

US-09-796-753-114

; Sequence 114, Application US/09796753
 ; Publication No. US20030027998A1
 ; GENERAL INFORMATION:
 ; APPLICANT: McCarthy, Sean A.
 ; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
 ; FILE REFERENCE: 7853-227-999
 ; CURRENT APPLICATION NUMBER: US/09/796,753
 ; CURRENT FILING DATE: 2001-03-01
 ; PRIOR APPLICATION NUMBER: 09/183,175
 ; PRIOR FILING DATE: 1998-10-30
 ; PRIOR APPLICATION NUMBER: 09/223,094
 ; PRIOR FILING DATE: 1998-12-30
 ; PRIOR APPLICATION NUMBER: 09/223,546
 ; PRIOR FILING DATE: 1998-12-30
 ; PRIOR APPLICATION NUMBER: 09/224,246
 ; PRIOR FILING DATE: 1998-12-30
 ; PRIOR APPLICATION NUMBER: 09/259,388
 ; PRIOR FILING DATE: 1999-02-26
 ; PRIOR APPLICATION NUMBER: 60/122,458
 ; PRIOR FILING DATE: 1999-03-01
 ; PRIOR APPLICATION NUMBER: 09/312,359
 ; PRIOR FILING DATE: 1999-05-14
 ; PRIOR APPLICATION NUMBER: 09/336,536
 ; PRIOR FILING DATE: 1999-06-18
 ; PRIOR APPLICATION NUMBER: 09/342,687
 ; PRIOR FILING DATE: 1999-06-29
 ; PRIOR APPLICATION NUMBER: 09/345,464
 ; PRIOR FILING DATE: 1999-06-30
 ; PRIOR APPLICATION NUMBER: 09/365,164
 ; PRIOR FILING DATE: 1999-07-30
 ; PRIOR APPLICATION NUMBER: 09/399,723
 ; PRIOR FILING DATE: 1999-09-20
 ; PRIOR APPLICATION NUMBER: 09/409,634
 ; PRIOR FILING DATE: 1999-09-30
 ; PRIOR APPLICATION NUMBER: 09/471,179
 ; PRIOR FILING DATE: 1999-12-23
 ; PRIOR APPLICATION NUMBER: 09/474,071

```

; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/474,072
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/514,010
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: 09/516,745
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/572,002
; PRIOR FILING DATE: 2000-05-14
; PRIOR APPLICATION NUMBER: 09/597,993
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 09/599,596
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/630,334
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/606,565
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/606,317
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/665,666
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 114
; LENGTH: 1050
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-753-114

```

```

Query Match          46.3%; Score 1667.5; DB 10; Length 1050;
Best Local Similarity 45.0%; Pred. No. 3.5e-103;
Matches 284; Conservative 61; Mismatches 181; Indels 105; Gaps 8;

```

```

Qy      14 LLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSCTDILNW----- 66
      ||| : | || ||| || |||:: | :||: || : | |
Db      9 LLLAVGLRLAGTLNPSDPNTCSFWESFTTTTKESHSRPFSLLPSEPCE--RPWEGPHTCP 66

Qy      67 -----FKCTRHRVSYSR-----TAY 80
      | | :|
Db      67 SPQTQRKLLASRDSFCMVCVGAGVQWRDRSALQPQTGNALSMRPQPRVLSGAPSLASPGH 126

Qy      81 RHGEKTMYSRRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGWGGTNCSSA-- 138
      || :||: ||| ||||| ||| || :||| ||| ||| ||| ||| :|||
Db      127 TVVVKTDHRQRLQCCHGFYESRGFCVPLCAQECVHGRCVAPNQCCVPGWRGDDCSSAPN 186

Qy      139 ----CDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGNDCHQR 194
      | :|| | |||| :|| | :| |||| | | | : | |||
Db      187 CLQPCTPGYYGPACQFRCQC-HGAPCDPQTGACFCPAERTGPSCDVSCSQGT----- 237

Qy      195 CQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGHPQCEQRCPCQNGGVCHHVTGECS 254
      :| || | | | ||||| | ||
Db      238 -----SGFFC-----PSTH-----PCQNGGVFQTPQGSCS 262

Qy      255 CPSGWMGTVCQGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSPGYTGERCQDECPVG 314
      || |||||:| |||| | |||||:|||| || ||| |:|||||:|:|:|||||
Db      263 CPPGWMGTICSLPCPEGFHGPNCSQECRCHNGGLCDREFTGQCRCAPGYTGDRCREECPPVG 322

```

Qy 315 TYGVLCAETCQCVNGGKCYHVSGACLCEAGFAGERCEARLCPEGLYGIKCDKRCPCCHLEN 374
 :| ||||| | :| :||| || |:|| |||:| ||: | | | :
 Db 323 RFGQDCAETCDCAPDARCFPANGACLCEHGF'TGDRCTDRLCPDGFYGLSCQAPCTCDREH 382
 Qy 375 THSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCDSVTGKCTCAPGF 434
 : |||||:||||:| |||:||:||||:| :| ||: | | :| | : :| | ||||:
 Db 383 SLSCHPMNGECSCCLPGWAGLHCNESCPODTHGPGCQEHCLCLHGGVCQATSGLCQCAPGY 442
 Qy 435 KGIDCSTPCPLGTYGINCSSRCGCKNDVAVCSFVDGSCCTKAGWHGVDCSIRCPSGTWGFG 494
 | |:: || |||:||:|| |::| ||:| | | | | :||: || |||||
 Db 443 TGPHCASLCPDPTYGVNCSARCSCEAIAACSPIDGECVCKEGWQRGNCSVPCPPGTWGF 502
 Qy 495 CNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAERCDCSHADGCHPTT 554
 || :||| : | : | ||| ||| | | :||| | :| || |||| | :||| |
 Db 503 CNASCQCAHEAVCSPQTGACTCTPGWHGAHCQLPCPKGQFGEGCASRCDCHSDGCDPVH 562
 Qy 555 GHCRCPLPGWSGVHCDVCAEGRWGPNCSLPC 585
 | |::| || | | | || || || |
 Db 563 GRCQCQAGWMGARCHLSCEPLWGVNCSNTC 593

RESULT 14

US-10-312-352-21

; Sequence 21, Application US/10312352

; Publication No. US20040053824A1

; GENERAL INFORMATION:

; APPLICANT: INCYTE GENOMICS, INC.; TANG, Y. Tom

; APPLICANT: YUE, Henry; AZIMZAI, Yalda

; APPLICANT: HE, Ann; BATRA, Sajeev

; APPLICANT: LO, Terence P.; NGUYEN, Dannel B.

; APPLICANT: BURRILL, John D.; MARCUS, Gregory A.

; APPLICANT: ZINGLER, Kurt A.; GANDHI, Ameena R.

; APPLICANT: LAL, Preeti G.; KEARNEY, Liam

; APPLICANT: BURFORD, Neil; YAO, Monique G.

; APPLICANT: CHAWLA, Narinder K.; ELLIOT, Vicki S.

; APPLICANT: ARVIZU, Chandra S.; KHAN, Farrah A.

; APPLICANT: BAUGHN, Mariah R.; HAFALIA, April, J.A.

; APPLICANT: POLICKY, Jennifer L.; AU-YOUNG, Janice K.

; APPLICANT: LU, Yan; BOROWSKY, Mark L.

; APPLICANT: LU, Dyung Aina M.; RAMKUMAR, Jayalaxmi

; APPLICANT: YANG, Junming; GURURAJAN, Rajagopal

; APPLICANT: WARREN, Bridget A.; GIETZEN, Kimberly J.

; APPLICANT: XU, Yuming; KALLICK, Deborah A.

; APPLICANT: LEE, Ernestine A.; THANGAVELU, Kavitha

; APPLICANT: DELEGEANE, Angelo M.; LEE, Sally

; TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES

; FILE REFERENCE: PF-0794 USN

; CURRENT APPLICATION NUMBER: US/10/312,352

; CURRENT FILING DATE: 2002-12-18

; PRIOR APPLICATION NUMBER: PCT/US01/21067

; PRIOR FILING DATE: 2001-06-29

; PRIOR APPLICATION NUMBER: US 60/215,454

; PRIOR FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: US 60/219,462

; PRIOR FILING DATE: 2000-07-18

; PRIOR APPLICATION NUMBER: US 60/240,111

```
; PRIOR FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: US 60/240,106
; PRIOR FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: US 60/244,021
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: US 60/248,887
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: US 60/249,570
; PRIOR FILING DATE: 2000-11-16
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PERL Program
; SEQ ID NO 21
; LENGTH: 1393
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20040053824A1 3351332CD1
US-10-312-352-21
```

```
Query Match          37.7%; Score 1356.5; DB 12; Length 1393;
Best Local Similarity 42.6%; Pred. No. 2.5e-82;
Matches 232; Conservative 50; Mismatches 207; Indels 56; Gaps 10;
```

```
Qy      92 SQC---CP-GFYESGEMCVPHCADKCVH-GRC-IAPNTCQCEPGWGGTNCSSACDGDHWG 145
      |:|  || |:|  | |  |:  |:  | |  | |  ||| |:|  |||  |||
Db      706 SRCQDVCPAGWY--GPSQCQTRCS--CANDGHCHPATGHCSCAPGWTGFSCQACDTGHWG 761

Qy      146 PHCTSRCQCKNG-ALCNPITGACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCQNGATCD 204
      | |:  | |  | |  |: |:| | | ||: | ||| |: | | |:| | | ||| |:| | |
Db      762 PDCSHPCNCSAGHGSCDAISGLCLCEAGYVGPRCEQQCPQGHFGPGCEQLCQCQHGAAACD 821

Qy      205 HVTGECRCPPGYTGAFCEDLCPGKHGPPQCEQRCPCQNGGVCHHVTGECSCPSGWMGTVC 264
      ||:| | || |: | ||| || | | | | | | | | | | | | | | | | | |
Db      822 HVSGACTCPAGWRGTFCEHACPAGEFFGLDCRSACNCTAGAACDAVNGSCLCPAGRRGPRC 881

Qy      265 GQPCPEGREFGKNCSQECQCHNGGTCDAAATGQCHCSP----- 300
      : ||  : | ||| | | || :||  ||||:|
Db      882 AETCPAHTYGHNCSSQACACFNGASCDPVHGGQCHCAPGWMGPSCLQECLPRDVRAGCRHSG 941

Qy      301 -----GYTGERCQDECPVGTYGVLCAETCQCVNGGKCYHVSGACL 341
      |:|:|:| | | |: | | | | | | | | | | | | | | | |
Db      942 GCLNGGLCDPHTGRCLCPAGWTGDKCQSPCLRGWFGEACAQRCSCPPGAACHHVTGACRC 1001

Qy      342 EAGFAGERCEARLCPEGLYGIKCDKRCPCCHLENTHSCHPMSGECACKPGWSGLYCNETCS 401
      || |  || : || | |: | : | | || :||| |: | |:| | : | | : |
Db      1002 PPGFTGSGCE-QACPPGSFGEDCAQMCQCPGENP-ACHPATGTCSAAGYHGPSCQQRC 1059

Qy      402 PGFYGEACQQICSCQNGADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKND 461
      || || |:|:| | || ||: || | | || | ||: || | |:| |:|  |||
Db      1060 PGRYPGCEQLCGCLNGGSCDAATGACRCPTGFLGTDNLTCPPQGRFGPNCTHVCGCGQG 1119

Qy      462 AVCSPVDGSCTCKAGWHGVDCSIRCPSGTWGFGCNLTCQCLNGGACNTLDGTCTCAPGWR 521
      | | || |:| | | || | ||  :| || | | ||| |:  |:|:|:|  ||
Db      1120 AACDPVTGTCLCPPGRAGVRCERGCPQNRFGVGCEHTCSCRNGGLCHASNGSCSCGLGWT 1179

Qy      522 GEKCELPQDGTGYGLNCAERCDCSHADGCHPTTGHCRLPGWSGVHCDSVCAEGRWGPNC 581
```

```

      | | | | | | | | | | : | | | | | | | : | | : | | | |
Db      1180 GRHCELACPPGRYGAACHLECSCHNNSTCEPATGTCRCGPGFYGQACEHPCPPGFGHAGC 1239

Qy      582 SLPCY 586
      | :
Db      1240 QGLCW 1244

```

RESULT 15

```

US-10-369-493-5280
; Sequence 5280, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION
OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5280
; LENGTH: 1111
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5280

```

```

Query Match          35.7%; Score 1284.5; DB 15; Length 1111;
Best Local Similarity 34.2%; Pred. No. 1.3e-77;
Matches 246; Conservative 77; Mismatches 221; Indels 175; Gaps 20;

```

```

Qy      21 GTASPLNLEDPNVCSHWESYSVTQESYPHPFDQIYYT-----SCTDILNWFKCTRHR 73
      | | | | : : | | : : | : : | | : | | : |
Db      35 GTTEP---QGDHVCT-----VKTIVDDY--ELKKVIHTVVYNDTEQCLNPLTGFQC---- 80

Qy      74 VSYRTAYRHGEKTMYYRRK-----SQCCPGFYESGE-MCVPHCADKCVHGRCIAPNTC 124
      | : | : | | : : | | | | : : : : | : | | | : | | |
Db      81 ----TVEKRGQKASYQRQLVKKEKYVKQCCDGYQTKDHFCLPDCNPPCKKKGKCIPEGKC 136

Qy      125 QCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRC----- 179
      : | : | : | | | : | : | | | : | | | : | : | | |
Db      137 ECDPGYGGKYCASSCSVGTWGLGCSKSCDCENGANCDPELGTCICTSGFQGERCEKPCPD 196

Qy      180 -----DRCEQGTYGNDCHQRCQCNQNGAT 202
      : | : | : | : | : | | | | |
Db      197 NKWGPNCVKSCPCQNGGKCNKEGKCVCSGDWGGGEFCLNKCEEKFGAECKFECNCQNGAT 256

Qy      203 CDHVTGECRCPPGYTGAFCEDLCPGKHPQCEQRCPCQNGGVCHHVTGECSCPSGWMGT 262
      | | : | : | | | | | : | | | | : | : | | | |
Db      257 CDNTNGKCICKSGYHGALCENECVSGFFGSGCTQKCDCLNNQNCDSGSECKC-IGWTGK 315

```


Qy 263 VCGQPCPEGRFGKNCSQECQC-----HNGGTCDAAATGQCHCSPGYTGERCQD-ECPVGT 315
 | | | | | | | | : : | | | | | | : : |
 Db 316 HCDIGCSRGRFGLQCKQNCTCPGLEFSDSNASCDAKTGQCQCESGYKGPCKDERKCDAEQ 375
 Qy 316 YGVLC AETCQCV--NGGKCYHVS GACLCEAGFAGERCEARLCPEGLYGIKCDKRCPC HLE 373
 | | : : | | | | : | | : | | : | | : | | : : |
 Db 376 YGADCSKTCTCVRENTLMCAPNTGFCRCKPGFYGDNCEL-ACSKDSYGPNCEKQAMCDWN 434
 Qy 374 NTHSCHPMMSGECACKPGW SGLYCNETCSPGFYGEACQQICSC-QNGADCDSVTGKCTCAP 432
 : | : | : | | | | : | | | | | | | | | |
 Db 435 HASECNPETGSCVCKPGRTGKNCSEPCPLDFYGPNAHQCCNQ RGVGCDGADGKCQCDR 494
 Qy 433 GFKGIDCSTPCPLGTYGINCSSRCGCKND AVCSPVDGSCTCKAGWHGVDCSIRCPSGTWG 492
 | : | | | | : | | | | | | : | | | | : : |
 Db 495 GWTGHRCEHHCPADTFGANCEKRCKCPKGIGCDPITGECTCPAGLQGANCDIGCPEGSYG 554
 Qy 493 FGCNLTCQCLNGGACNTLDGTCTCAPGW-----RGEKCEL--PCQD----- 531
 | | | : : | | : | | | | : | | | | | |
 Db 555 PGCKLHCKCVN-GKCDKETGECTCQPGFFGSDCSTTCSKGKYGESCELS CPCSDASCSKQ 613
 Qy 532 -----GTYGLNCAERCD----- 543
 | | : : | : : |
 Db 614 TGKCLCPLGTKGVSCDQKCDPNTFGFLCQETVTPSPCASTDPKNGVCLSCPPGSSGIHCE 673
 Qy 544 -----CSHAD--GCHPTTGHCRCPLGW SGVHCD SVCAEGRWGPNC SLPC 585
 | | | | | | | : | | | : : | : : |
 Db 674 HNC PAGSYGDGCQQVCSCADGHGCDPTTGECICEPGYHGKTCSEKCPDGKYGYGCALDC 732

Search completed: March 26, 2004, 16:21:16
 Job time : 28.1611 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2004, 16:04:46 ; Search time 28.5191 Seconds
(without alignments)
6483.148 Million cell updates/sec

Title: US-10-092-390-4
Perfect score: 3601
Sequence: 1 MVISLNSCLSFCILLLCHWI.....HCDSVCAEGRWGPNCSLPCY 586

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertibrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

%
Result Query
No. Score Match Length DB ID Description

1	3601	100.0	1140	4	Q96KG7	Q96kg7 homo sapien
2	3468	96.3	567	4	Q8WUL3	Q8wul3 homo sapien
3	2343.5	65.1	947	11	Q8BKK7	Q8bkk7 mus musculu
4	2094	58.2	969	4	Q96KG6	Q96kg6 homo sapien
5	1828	50.8	747	11	Q8VHF4	Q8vhf4 mus musculu
6	1828	50.8	1034	11	Q8VHL7	Q8vhl7 mus musculu
7	1824	50.7	1034	11	Q8VIK5	Q8vik5 mus musculu
8	1690	46.9	1004	11	Q8CGA7	Q8cga7 mus musculu
9	1575	43.7	921	11	Q80T91	Q80t91 mus musculu
10	1372.5	38.1	1574	11	O88281	O88281 rattus norv
11	1362	37.8	626	4	Q8ND91	Q8nd91 homo sapien
12	1341	37.2	299	11	Q8BX64	Q8bx64 mus musculu
13	1340	37.2	1664	5	Q9TVQ2	Q9tvq2 caenorhabdi
14	1309	36.4	220	11	Q63404	Q63404 rattus norv
15	1290.5	35.8	881	5	Q9W0A0	Q9w0a0 drosophila
16	1284.5	35.7	1045	5	Q8T3A6	Q8t3a6 caenorhabdi
17	1284.5	35.7	1070	5	Q8T3A7	Q8t3a7 caenorhabdi
18	1284.5	35.7	1111	5	Q9XWD6	Q9xwd6 caenorhabdi
19	1282.5	35.6	1246	4	O75095	O75095 homo sapien
20	1230.5	34.2	546	11	Q80V70	Q80v70 mus musculu
21	843.5	23.4	569	4	Q8NHD4	Q8nhd4 homo sapien
22	839	23.3	320	4	Q8N780	Q8n780 homo sapien
23	800	22.2	866	4	Q8IXF3	Q8ixf3 homo sapien
24	741	20.6	594	5	Q9W0A1	Q9w0a1 drosophila
25	740	20.5	594	5	Q9Y151	Q9y151 drosophila
26	708.5	19.7	2447	13	O13149	O13149 fugu rubrip
27	702.5	19.5	337	4	Q8NHD3	Q8nhd3 homo sapien
28	702.5	19.5	342	4	Q8NHD5	Q8nhd5 homo sapien
29	693	19.2	2516	11	Q7TQ52	Q7tq52 mus musculu
30	693	19.2	2526	11	Q7TQ51	Q7tq51 mus musculu
31	693	19.2	2531	11	Q8K428	Q8k428 mus musculu
32	693	19.2	2531	11	Q7TQ50	Q7tq50 mus musculu
33	687	19.1	744	4	Q8NHD2	Q8nhd2 homo sapien
34	685.5	19.0	2524	5	Q9GPA5	Q9gpa5 branchiosto
35	682.5	19.0	2468	13	Q800E4	Q800e4 brachydanio
36	682	18.9	4288	4	Q9NPK9	Q9npk9 homo sapien
37	681.5	18.9	2531	5	O16004	O16004 lytechinus
38	681	18.9	2653	5	Q25253	Q25253 lucilia cup
39	678	18.8	1193	13	Q90819	Q90819 gallus gall
40	676.5	18.8	4006	11	O35452	O35452 mus musculu
41	672.5	18.7	4135	6	O18977	O18977 bos taurus
42	667.5	18.5	4114	11	O54796	O54796 mus musculu
43	666	18.5	2528	13	Q8AXP0	Q8axp0 cynops pyrr
44	664	18.4	1214	13	Q90YD2	Q90yd2 xenopus lae
45	657.5	18.3	2428	5	Q8I6X6	Q8i6x6 boophilus m

ALIGNMENTS

RESULT 1

Q96KG7

ID	Q96KG7	PRELIMINARY;	PRT;	1140 AA.
AC	Q96KG7;			
DT	01-DEC-2001	(TrEMBLrel. 19, Created)		
DT	01-DEC-2001	(TrEMBLrel. 19, Last sequence update)		
DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)		

DE MEGF10 protein (Hypothetical protein KIAA1780).
 GN MEGF10.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Hippocampus;
 RX MEDLINE=21245130; PubMed=11347906;
 RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XX.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large Proteins in vitro.";
 RL DNA Res. 8:85-95(2001).
 DR EMBL; AB058676; BAB47409.1; -.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR009030; Grow_fac_recep.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00008; EGF; 10.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00180; EGF_Lam; 6.
 DR PROSITE; PS00022; EGF_1; 17.
 DR PROSITE; PS01186; EGF_2; 17.
 KW Hypothetical protein; EGF-like domain; Laminin EGF-like domain.
 SQ SEQUENCE 1140 AA; 122204 MW; 45B2FA239423895A CRC64;

Query Match 100.0%; Score 3601; DB 4; Length 1140;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 586; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MVISLNSCLSFICLLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC	60
Db	1	MVISLNSCLSFICLLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC	60
Qy	61	TDILNWFKCTRHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIA	120
Db	61	TDILNWFKCTRHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIA	120
Qy	121	PNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCED	180
Db	121	PNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCED	180
Qy	181	RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKGHPQCEQRCPC	240
Db	181	RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKGHPQCEQRCPC	240
Qy	241	QNGGVCHHVTGECSCPSGWMGTVCGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP	300
Db	241	QNGGVCHHVTGECSCPSGWMGTVCGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP	300
Qy	301	GYTGERCQDECPVGTYGVLCAETCQCVNNGKCYHVSAGLCEAGFAGERCEARLCPEGLY	360
Db	301	GYTGERCQDECPVGTYGVLCAETCQCVNNGKCYHVSAGLCEAGFAGERCEARLCPEGLY	360
Qy	361	GIKCDKRCPCHLENTHSCHPMSEGCACKPGWGLYCNETCSPGFYGEACQQICSCQNGAD	420

```

Db          361 GIKCDKRCPCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGAD 420
Qy          421 CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSPVDGSCTCKAGWHGV 480
Db          421 CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSPVDGSCTCKAGWHGV 480
Qy          481 DCSIRCPSGTWGFGCNLTQCCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTGYGLNCAE 540
Db          481 DCSIRCPSGTWGFGCNLTQCCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTGYGLNCAE 540
Qy          541 RCDCSHADGCHPTTGHCRCPLPGWSGVHCDSVCAEGRWGPNCSLPCY 586
Db          541 RCDCSHADGCHPTTGHCRCPLPGWSGVHCDSVCAEGRWGPNCSLPCY 586

```

RESULT 2

Q8WUL3

```

ID   Q8WUL3          PRELIMINARY;          PRT;    567 AA.
AC   Q8WUL3;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to MEGF10 protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Muscle;
RA   Strausberg R.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC020198; AAH20198.1; -.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002049; Laminin_EGF.
DR   Pfam; PF00008; EGF; 7.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00180; EGF_Lam; 4.
DR   PROSITE; PS00022; EGF_1; 10.
DR   PROSITE; PS01186; EGF_2; 10.
KW   EGF-like domain; Laminin EGF-like domain.
SQ   SEQUENCE    567 AA;  60797 MW;  CF2FB8CDEB7CF627 CRC64;

```

```

Query Match          96.3%;  Score 3468;  DB 4;  Length 567;
Best Local Similarity 99.8%;  Pred. No. 0;
Matches 565;  Conservative 1;  Mismatches 0;  Indels 0;  Gaps 0;

```

```

Qy          1 MVISLNSCLSFICLLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC 60
Db          1 MVISLNSCLSFICLLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC 60
Qy          61 TDILNWFKCTRHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIA 120
Db          61 TDILNWFKCTRHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIA 120

```

Qy	121	PNTCQCEPGWGGTNCSSACGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCED	180
Db	121	PNTCQCEPGWGGTNCSSACGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCED	180
Qy	181	RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPPQCEQRCPC	240
Db	181	RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPPQCEQRCPC	240
Qy	241	QNGGVCHHVTGECSCPSGWMGTVCQGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP	300
Db	241	QNGGVCHHVTGECSCPSGWMGTVCQGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP	300
Qy	301	GYTGERCQDECPVGTYGVLCAETCQCVNGGKCYHVSAGLCEAGFAGERCEARLCPEGLY	360
Db	301	GYTGERCQDECPVGTYGVLCAETCQCVNGGKCYHVSAGLCEAGFAGERCEARLCPEGLY	360
Qy	361	GIKCDKRCPCCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGAD	420
Db	361	GIKCDKRCPCCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGAD	420
Qy	421	CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSPVDGSCTCKAGWHGV	480
Db	421	CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSPVDGSCTCKAGWHGV	480
Qy	481	DCSIRCPSGTWGFGCNLTQCCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAE	540
Db	481	DCSIRCPSGTWGFGCNLTQCCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAE	540
Qy	541	RCDCSHADGCHPTTGHCRCPLPGWSGV	566
Db	541	RCDCSHADGCHPTTGHCRCPLPGWSGL	566

RESULT 3

Q8BKK7

ID Q8BKK7 PRELIMINARY; PRT; 947 AA.
AC Q8BKK7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MEGF11 protein.
GN 2410080H04RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Dorsal root ganglion;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK051642; BAC34702.1; -.
DR MGD; MGI:1920951; 2410080H04Rik.

DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0045285; C:ubiquinol-cytochrome-c reductase complex; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0008121; F:ubiquinol-cytochrome-c reductase activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR005805; Rieske.
 DR Pfam; PF00008; EGF; 11.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00181; EGF; 15.
 DR SMART; SM00180; EGF_Lam; 15.
 DR PROSITE; PS00022; EGF_1; 15.
 DR PROSITE; PS01186; EGF_2; 15.
 DR PROSITE; PS00200; RIESKE_2; 1.
 SQ SEQUENCE 947 AA; 100661 MW; 0C209B11DFEE8314 CRC64;

Query Match 65.1%; Score 2343.5; DB 11; Length 947;
 Best Local Similarity 63.7%; Pred. No. 1.2e-215;
 Matches 364; Conservative 55; Mismatches 121; Indels 31; Gaps 1;

Qy	15	LLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSCTDILNWFKCTRHRV	74
		: : :	
Db	8	LLVFLQLAALALNPEDPNVCSHWESYAVTVQESYAHFPDQIYYTRCADILNWFKCTRHRH	67
Qy	75	SYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGWGGTN	134
		: : : : : : :	
Db	68	SYKTAYRRGLRTMYRRRSQCCPGYYENGDFCI-----	99
Qy	135	CSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGNDCHQR	194
		: : : : : :	
Db	100	---RCDSEHWGPHCSNRCQCNQNGALCNPITGACVCAPGFRGWRCEELCAPGTHGKGCQLL	156
Qy	195	CQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPQCEQRCPCQNGGVCHHVTGEC	254
		: : : : : :	
Db	157	CQCHHGASCDPRTGECCLCAPGYTGVCYCEELCPPGSHGAHCELRCPQNGGTCHHITGECA	216
Qy	255	CPSGWMGTVCQGPCPEGRFGKNCSEQECQCHNGGTCDAAATGQCHCSPGYTGERCQDECPVG	314
		: : : : : :	
Db	217	CPPGWTGAVCAQPCPPGTFGQNCSDCPCHHGGQCDHVTGQCHCTAGYMGDRQCQEECPFG	276
Qy	315	TYGVLCAETCQCVNGGKCYHVSGACLCEAGFAGERCEARLCPEGLYGIKCDKRCPCCHLEN	374
		: : : : : : :	
Db	277	TFGFLCSQRCDCHNGGQCSPATGACECEPGYKGPSCQERLCPEGLHGPCTLPCCPDEN	336
Qy	375	THSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCDSVTGKCTCAPGF	434
		: : : : : :	
Db	337	TISCHPVTGACTCQPGWSGHYCNESCPAGYYNGCQLPCTCQNGADCHSITGSCTCAPGF	396
Qy	435	KGIDCSTPCPLGTYGINCSSRCGCKNDAVCSVPDGSCTCKAGWHGVDCSIRCPSGTWGFG	494
		: : : :	
Db	397	MGEVCAVPCAAGTYGPNCSVCSNNGGTCSVPDGSCTCREGWQLDCSLPCPSGTWGLN	456
Qy	495	CNLTQCQLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAERCDCSHADGCHPTT	554
		: : : :	
Db	457	CNETCICANGAACSPFDGSCACTPGWLGDSCELPCPDGTGFLNCSEHCDCSHADGCDPVT	516

Qy 555 GHCRCPLPGWSGVHCDSVCAEGRWGPNC SLPC 585
 ||| || ||:|: ||| | |||||: |
 Db 517 GHCCCLAGWTGIRCDSTCPPGRWGPNC SVSC 547

RESULT 4

Q96KG6

ID Q96KG6 PRELIMINARY; PRT; 969 AA.
 AC Q96KG6;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE MEGF11 protein (Hypothetical protein KIAA1781).
 GN MEGF11.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=21245130; PubMed=11347906;
 RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XX.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large Proteins in vitro.";
 RL DNA Res. 8:85-95(2001).
 DR EMBL; AB058677; BAB47410.1; -.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0045285; C:ubiquinol-cytochrome-c reductase complex; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0008121; F:ubiquinol-cytochrome-c reductase activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR005805; Rieske.
 DR Pfam; PF00008; EGF; 12.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00180; EGF_Lam; 8.
 DR PROSITE; PS00022; EGF_1; 17.
 DR PROSITE; PS01186; EGF_2; 17.
 DR PROSITE; PS00200; RIESKE_2; 1.
 KW Hypothetical protein; EGF-like domain; Laminin EGF-like domain.
 SQ SEQUENCE 969 AA; 101600 MW; 56DD2FFE139C8209 CRC64;

Query Match 58.2%; Score 2094; DB 4; Length 969;
 Best Local Similarity 65.4%; Pred. No. 9.1e-192;
 Matches 312; Conservative 58; Mismatches 107; Indels 0; Gaps 0;

Qy 109 CADKCVHGRCIAPNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACH 168
 | :|||||:|:| | ||||| :|| | |||||:|:|:| |||||
 Db 28 CTEECVHGRCVSPDTCHEPGWGGPDCSSGCSDSHWGPHCNRCQCNALCNPITGACV 87
 Qy 169 CAAGFRGWRCEDRCEQGTGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCP PG 228
 |||||:| | ||:| | ||:|:| | || | |||:|:|
 Db 88 CAAGFRGWRCEELCAPGTHGKGCQLPCQCRHGASCDPRAGECLCAPGYTGVYCEELCP PG 147

QY 229 KHGPQCEQRCPCQNGGVCHHVTGECSCPSGWMGTVCQPCPEGRFGKNCSQECQCHNGGT 288
 || || ||||| |||:||||:| || || || || |||:| ||:|
 Db 148 SHGAHCELRCPCQNGGTCHHITGECACPPGWTGAVCAQPCPPGTFGQNCSQDCPCHHGGQ 207
 QY 289 CDAATGQCHCSPGYTGERCQDECPVGTYGVLCAETCQCVNGGKCYHVSAGACLCEAGFAGE 348
 || |||||: || |:|:| || |:| |:| || |:| |:| || |:|
 Db 208 CDHVTGQCHCTAGYMGDRCEECFPGSFGFQCSQRCDCHNGGQCSPTTGACECEPGYKGP 267
 QY 349 RCEARLCPEGLYGIKCDKRCPCCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEA 408
 ||: |||||:| | || |:| ||||:| |:| || |:| |:| |:|
 Db 268 RCQERLCPEGLHGPCTLPCCDADNTISCHPVTGACTCQPGWSGHHHCNESCPVGYGDG 327
 QY 409 CQQICSCQNGADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDVAVCSPVD 468
 || |:| ||||| |:| ||||| | |:| | ||| |||| | | |||||
 Db 328 CQLPCTCQNGADCHSITGGCTCAPGFMGEVCAVSCAAGTYGPNCSSICSCNNGGTCSPVD 387
 QY 469 GSCTCKAGWHGVDCSIRCPSGTWGFGCNLTQCCLNGGACNTLDGTCTCAPGWRGEKCELP 528
 ||||| || |:|:| ||||| || |:| || || |:| |:|:| ||| |:| |||
 Db 388 GSCTCKEGWQGLDCTLPSPGTWGLNCNESCTCANGAACSPIDGSCSCTPGWLGDTCCELP 447
 QY 529 CQDGTGYNLCAERCDCSHADGCHPTTGHCRCPLPGWSGVHCDVCAEGRWGPNCSLPC 585
 | |||:||||:| ||||| || |||| || |:|:| ||| | |||||:|
 Db 448 CPDGTGFLNCSEHCDCSHADGCDPVTGHCCCLAGWTGIRCDSTCPPGRWGPNCVSC 504

RESULT 5

Q8VHF4

ID Q8VHF4 PRELIMINARY; PRT; 747 AA.
 AC Q8VHF4;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Jedi-736 protein.
 GN 3110045G13RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL; TISSUE=Testis;
 RA Krivtsov A.V., Zinovyeva M.V., Hendrikx J., Visser J.W.M.,
 RA Belyavsky A.V.;
 RT "Jedi is a novel DSL and EGF-like repeat motif-containing protein
 RT expressed on non-differentiated hematopoietic cells.";
 RL Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AF461685; AAL66380.1; -.
 DR MGD; MGI:1920432; 3110045G13Rik.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR009030; Grow_fac_recep.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00008; EGF; 6.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00180; EGF_Lam; 4.
 DR PROSITE; PS00022; EGF_1; 13.

DR PROSITE; PS01186; EGF_2; 12.
KW EGF-like domain; Laminin EGF-like domain.
SQ SEQUENCE 747 AA; 78972 MW; F825F8F384D4736A CRC64;

Query Match 50.8%; Score 1828; DB 11; Length 747;
Best Local Similarity 52.0%; Pred. No. 2e-166;
Matches 299; Conservative 59; Mismatches 211; Indels 6; Gaps 3;

```
Qy      14 LLLCHWIGTASPLNLEDPNVC SHWESYSVTVQESYPHPFDQIYYTSCTDILNW---FKCT 70
      ||| : || |||||: |||: | :||: || : || |
Db      7 LLLALGLRLTGTLNSNDPNVCTFWESFTTTTKESHLRPFSLPAESCH--RPWEDPHTCA 64

Qy      71 RHRVSYRTAYRHGEKTM YRRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGW 130
      : | ||| || | : ||| :||| ||| || :||| ||| ||| |||
Db      65 QPTVVYRTVYRQVVKMDSRPLQCCRGYYESRGACVPLCAQECVHGRCVAPNQCQCAGW 124

Qy     131 GGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGND 190
      | :||| | ||| | | | : |:| :||| | :| : | | ||
Db     125 RGGDCSSECAPGMWGPQCDKFCHCGNNSSCDPKSGACFCPSGLQPPNCLQPCPAGHYGPA 184

Qy     191 CHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPPQCEQRCPCQNGGVCHHVT 250
      | ||| ||:| | |||| | | | | | | | | | | | | | |
Db     185 CQFDCQCY-GASCDPQDGACFCPPGRAGPSCNVPCSQGTDGFFCPRTYPCQNGGVPGQSQ 243

Qy     251 GECSCPSGWMGTVCGQPCPEGREFGKNCSQECQCHNGGTCDAATGQCHCSPGYTGERCQDE 310
      | |||| |||| :| |||| | ||:| |:||| || |||||: ||| |:| |:|
Db     244 GSCSCPPGWMGVICSLPCPEGFHGPNTQECRCHNGGLCDRFTGQCHCAPGYIGDRCQEE 303

Qy     311 CPVGTYGVLCAETCQCVNGGKCYHVS GACLCEAGFAGERCEARLCPEGLYGIKCDKRCPC 370
      |||| :| |||| | | :|: :||| || | :|| ||||:| ||: | :| |
Db     304 CPVGRFGQDCAETCD CAPGARCFFPANGACLCEHGFTGDRCTERLCPDGRYGLSCQEPCTC 363

Qy     371 HLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCDSVTGKCTC 430
      | :| |||| |||:| |:| |:| |:| |:| | :| ||: | | :| | : :| | |
Db     364 DPEHSLSCHPMHGECSCQPGWAGLHCNESCPQDTHGPGCQEHCLCLHGGLCLADSGLCRC 423

Qy     431 APGFGKIDCSTPCPLGTYGINCSSRCGCKNDVCSVPVDGSCCTKAGWHGVDCSIRCPSTG 490
      |||: | | : || ||||| ||| |:| |||:| |:| || | :||: || ||
Db     424 APGYTGPHCANLCPDPTYGINCSSRCSCENAIACSPIDGTICKEGWQRGNCSVPCPLGT 483

Qy     491 WGFGCNLTQCCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAERCDCSHADGC 550
      ||| || :||| : | | : | ||| ||| | |:|| | :| || ||| |:|||
Db     484 WGFNCNASQCQCAHDGVCSPQTGACTCTPGWHGAHCQLPCPKGQFGEGCASVDCDHS DGC 543

Qy     551 HPTTGHCRCPLPGWSGVHCD SVCAEGRWGPNC SLPC 585
      | | ||| || | | | ||| ||| |
Db     544 DPVHGQCRCQAGWMGTRCHLPCPEGFWGANC SNTC 578
```

RESULT 6

Q8VHL7

ID Q8VHL7 PRELIMINARY; PRT; 1034 AA.

AC Q8VHL7;

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Jedi protein.

GN 3110045G13RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OK NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL; TISSUE=Testis;
 RA Krivtsov A.V., Zinovyeva M.V., Hendrikx J., Visser J.W.M.,
 RA Belyavsky A.V.;
 RT "Jedi is a novel DSL and EGF-like repeat motif-containing protein
 RT expressed on non-differentiated hematopoietic cells.";
 RL Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AF444274; AAL38571.1; -.
 DR MGD; MGI:1920432; 3110045G13Rik.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR009030; Grow_fac_recep.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00008; EGF; 6.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00180; EGF_Lam; 4.
 DR PROSITE; PS00022; EGF_1; 13.
 DR PROSITE; PS01186; EGF_2; 12.
 KW EGF-like domain; Laminin EGF-like domain.
 SQ SEQUENCE 1034 AA; 110540 MW; 5514E5166AE01111 CRC64;

Query Match 50.8%; Score 1828; DB 11; Length 1034;
 Best Local Similarity 52.0%; Pred. No. 2.8e-166;
 Matches 299; Conservative 59; Mismatches 211; Indels 6; Gaps 3;

Qy 14 LLLCHWIGTASPLNLEDPNVC SHWESYSVTVQESYPHPFDQIYYTSCTDI LNW---FKCT 70
 ||| : || |||||: |||: | :||: || : || |
 Db 7 LLLALGLRLTGT LNSNDPNVCTFWESFTTTTKESH LRPFSLLPAESCH--RPWEDPHTCA 64
 Qy 71 RHRVSYRTAYRHGEKTM YRRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGW 130
 : | ||| || | : ||| |:||| ||| || :||| ||| ||| ||| |||
 Db 65 QPTVVYRTVYRQVVKMSRPR LQCCRGYYESRGACVPLCAQECVHGRCVAPNQCCQCAPGW 124
 Qy 131 GGTNCSSACDGDHWGPHCTSR CQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGT YGND 190
 | :||| | ||| | | | : ||| :||| | :| : | | |||
 Db 125 RGGDCSSECAPGMWGPQCDKFCHCGNNSCDPKSGACFCPSGLQPPNCLQPCPAGHYGPA 184
 Qy 191 CHQRCQCQNGATCDHVTGECRCPPGYTGAFCE DLCP PGKHGPQCEQRCP CQNGGVCHHVT 250
 | ||| ||:|| | | |||| | | | | | : |||||
 Db 185 CQFDCQCY-GASCDPQDGACFCPPGRAGPSCNVPCSQGT DGF FCPRTYPCQNGGVPQGSQ 243
 Qy 251 GECSCPSGWMGTVC GQPCPEGRF GKNC SQECQCHNGGTCD AATGQCHCSPGYTGERCQDE 310
 | |||| |||| :| |||| | ||:||||:|||| || |||||:||| |:||:|
 Db 244 GSCSCPPGWMGVICSLPCPEGFHGP NCTQECRCHNGGLCDRFTGQCHCAPGYIGDR CQEE 303
 Qy 311 CPVGT YGVLCAETCQCVN GGKCYH VSGACLCEAGFAGERCEARLCPEGLYGIKCDKRCPC 370
 |||| :| |||| | | :| : ||||| || |:|| ||||:| ||: | : | |
 Db 304 CPVGRFGQDCAETCD CAPGARCFPANGACLCEHGFTGDRCTERLC PDGRYGLSCQEPCTC 363
 Qy 371 HLENTHSCHPMSGECACKPGW SGLYCNETCSPGFYGEACQ QICSCQNGADCD SVTGKCTC 430
 |:: |||| |||:|:|||:||:||||:| :| ||: | | :| | : :| | |

Db 364 DPEHSLSCHPMHGECSQPGWAGLHCNESCQDTHGPGCQEHCLCLHGGLCLADSGLCRC 423
 Qy 431 APGFKGIDCSTPCPLGTYGINCSSRCCKNDVCSFVDSCTCKAGWHGVDCSIRCPSGT 490
 |||:| |: || ||||| ||| |:| |||:|:| || || |:|: || ||
 Db 424 APGYTGPHCANLCPDITYGINCSSRCSCEAIACSPIDGTICKEGWQRGNCSVPCPLGT 483
 Qy 491 WGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTGYLNCARCDSCSHADGC 550
 ||| || |:| |:| |:| | ||| ||| | |:| | |:| || || |:|
 Db 484 WGFNCNASCQCAHDGVCSPTGACTCTPGWHGAHCQLPCPKGQFGEGCASVCDCHSDGC 543
 Qy 551 HPTTGHCRLPGWSGVHCDSVCAEGRWGPNCSLPC 585
 | | ||| || | | | ||| ||| |
 Db 544 DPVHGQCRCQAGWMGTRCHLPCPEGEFWGANCSNTC 578

RESULT 7

Q8VIK5

ID Q8VIK5 PRELIMINARY; PRT; 1034 AA.
 AC Q8VIK5;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE MEGF12.
 GN 3110045G13RIK OR MEGF12.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Liver;
 RA Ivanova N.B., Lemischka I.R.;
 RT "The global gene expression profiling of the hematopoietic stem
 cell.";
 RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Eye;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 DR EMBL; AF440279; AAL33583.1; -.
 DR EMBL; AK053551; BAC35426.1; -.
 DR MGD; MGI:1920432; 3110045G13Rik.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR009030; Grow_fac_recep.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00008; EGF; 6.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00180; EGF_Lam; 4.
 DR PROSITE; PS00022; EGF_1; 13.
 DR PROSITE; PS01186; EGF_2; 12.
 KW EGF-like domain; Laminin EGF-like domain.

SQ SEQUENCE 1034 AA; 110580 MW; 714E5016848E4E4C CRC64;

Query Match 50.7%; Score 1824; DB 11; Length 1034;
Best Local Similarity 51.8%; Pred. No. 6.7e-166;
Matches 298; Conservative 59; Mismatches 212; Indels 6; Gaps 3;

```
Qy      14 LLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSCTDILNW---FKCT 70
      ||| : || |||||: |||: | :||: || : || | |
Db      7 LLLALGLRLTGTLNSNDPNVCTFWESFTTTTKESHLRPFSLPAESCH--RPWEDPHTCA 64

Qy      71 RHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGW 130
      : | ||| || | : ||| |:||| ||| || :||| ||: ||| ||| |||
Db      65 QPTVVYRTVYRQVVKMDSRPLQCCRGYYESRGACVPLCAQECVHGRCVAPNQCQCAGW 124

Qy     131 GGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGND 190
      | :||| | ||| | | | : | :| :| | | : | | | |
Db     125 RGGDCSSECAPGMWGPQCDKFCHCGNNSCDPKSGTCFCPSGLQPPNCLQPCPAGHYGPA 184

Qy     191 CHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPQCEQRCPCQNGGVCHHVT 250
      | ||| ||: || | |||| | | | | | : |||||
Db     185 CQFDCQCY-GASCDPQDGACFCPPGRAGPSCNVPCSQGTDFGFCPRTPCQNGGVPQGSQ 243

Qy     251 GECSCPSGWMGTVCQGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSPGYTGERCQDE 310
      | |||| |||| :| |||| | ||: ||: |||| || |||||: ||| | :||| :|
Db     244 GSCSCPPGWMGVICSLPCPEGFHGPNECTQECCRCHNGGLCDRFTGQCHCAPGYIGDRCQEE 303

Qy     311 CPVGTYGVLCAETCQCVNGGKCYHVSAGLCEAGFAGERCEARLCPEGLYGIKCDKRCPC 370
      |||| :| |||| | | :| : ||||| || | :|| ||||: | ||: | : | |
Db     304 CPVGRFGQDCAETCDCAPGARCFPANGACLCEHGFTGDRCTERLCPDGRYGLSCQEPCTC 363

Qy     371 HLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCDSVTGKCTC 430
      |:: |||| |||: |||: |||: |||: | :| ||: | | :| | : :| | |
Db     364 DPEHSLSCHPMHGECSQPGWAGLHCNESCPQDTHGPGCQEHCLCLHGGGLCLADSGLCRC 423

Qy     431 APGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSFVDGSCCTCKAGWHGVDCSIRCPST 490
      |||: | |: || ||||| ||| |: | |||: ||: | || | :||: || ||
Db     424 AFGYTGPHCANLCPDPTYGINCSSRCSCEAIACSPIDGTICKEGWQRGNCSVPCPLGT 483

Qy     491 WGFGCNLTQCQLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAERCDCSHADGC 550
      ||| || :||| : | |: | ||| ||| | |: ||| | :| || ||| |: |||
Db     484 WGFNCNASQCQAHDGVCSPQTGACTCTPGWHGAHCQLPCPKGQFGEGCASVCDCHSDGC 543

Qy     551 HPTTGHCRCPLPGWSGVHCDVCAEGRWGPNCSLPC 585
      | | ||| || | | | ||| ||| || |
Db     544 DPVHGQCRCQAGWMGTRCHLPCPEGFWGANCSTC 578
```

RESULT 8

Q8CGA7

ID Q8CGA7 PRELIMINARY; PRT; 1004 AA.

AC Q8CGA7;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Similar to RIKEN cDNA 3110045G13 gene.

GN 3110045G13RIK.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N;
 RA Strausberg R.;
 RL Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; BC042490; AAH42490.1; -.
 DR MGD; MGI:1920432; 3110045G13Rik.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR009030; Grow_fac_recep.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00008; EGF; 6.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00181; EGF; 14.
 DR SMART; SM00180; EGF_Lam; 12.
 DR PROSITE; PS00022; EGF_1; 12.
 DR PROSITE; PS01186; EGF_2; 11.
 SQ SEQUENCE 1004 AA; 107377 MW; 9508B0EC04561E94 CRC64;

Query Match 46.9%; Score 1690; DB 11; Length 1004;
 Best Local Similarity 49.2%; Pred. No. 4.4e-153;
 Matches 283; Conservative 52; Mismatches 204; Indels 36; Gaps 4;

Qy	14	LLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSCTDILNW---FKCT	70
		: : : : :	
Db	7	LLLALGLRLTGTLNSNDENVCTFWESFTTTTKESHRLRPFSLLPAESCH--RPWEDPHTCA	64
Qy	71	RHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGW	130
		: : : :	
Db	65	QPTVVYRTVYRQVVKMDSRPRLQCCRGYYESRGACVPLCAQECVHGRCVAPNQCCAPGW	124
Qy	131	GGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGND	190
		: : : : : :	
Db	125	RGGDCSSECAPGMWGPQCDKFCHCGNNSCDPKSGACFCPSGLQPPNCLQPCPAGHYGPA	184
Qy	191	CHQRCQCQNGATCDHVTGECRCPPGYTGAFCELDLCPGKGHPQCEQRCPCQNGGVCHHVT	250
		: :	
Db	185	CQFDCQCY-GASCDPQDGACFCPPGRAGPSCNVPCSQGTGDGFFCPRTYPCQNGGVPQGSQ	243
Qy	251	GECSCPSGWMGTVCQGQPCPEGRFGKNCSEQCQCHNGGTCDAAATGQCHCSPGYTGERCQDE	310
		: : :	
Db	244	GSCSCPPGWMGVICSLPCPEGFHGPNCQTQECRCHNGGLCDRFTGQCHCAPGYIGDRCQEE	303
Qy	311	CPVGTYGVLC AETCQCVNGGKCYHVS GACLCEAGFAGERCEARLCPEGLYGIKCDKRCPC	370
		: : : : : : :	
Db	304	CPVGRFGQDCAETCDCAFGARCFPANGACLCEHGFTGDRCTERLCPDGRYGLSCQEPCTC	363
Qy	371	HLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCDSVTGKCTC	430
		: : : : :	
Db	364	DPEHSLSCHPMH-----CLCLHGGLCLADSGLCRC	393
Qy	431	APGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSPVDGSCCTCKAGWHGVDCSIRCPSGT	490
		: : : : : : :	

Db 394 APGYTGPHCANLCPDPTYGINCSSRCSCEAIACSPIDGTICKEGWQRGNCSVPCPLGT 453

QY 491 WGFGCNLTQCCLNGGACNTLDGTCTCAPGWRGEKCELPCQDGTYGLNCAERCDCSHADGC 550
 ||| || :||| : | | : | ||| ||| | | :||| | :| || ||| |:|||

Db 454 WGFNCNASCQCAHDGVCSPQTGACTCTPGWHGAHCQLPCPKGQFGEGCASVCDCHSDGC 513

QY 551 HPTTGHCRCPLPGWSGVHCDSVCAEGRWGPNCSLPC 585
 | | ||| || | | | ||| ||| |

Db 514 DPVHGQCRCQAGWMGTRCHLPCPEGFWGANCSNTC 548

RESULT 9

Q80T91

ID Q80T91 PRELIMINARY; PRT; 921 AA.

AC Q80T91;

DT 01-JUN-2003 (TrEMBLrel. 24, Created)

DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE MKIAA1781 protein (Fragment).

GN MKIAA1781.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=22579291; PubMed=12693553;

RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,

RA Nakajima D., Nagase T., Ohara O., Koga H.;

RT "Prediction of the coding sequences of mouse homologues of KIAA gene:

RT II. The complete nucleotide sequences of 400 mouse KIAA-homologous

RT cDNAs identified by screening of terminal sequences of cDNA clones

RT randomly sampled from size-fractionated libraries.";

RL DNA Res. 10:35-48(2003).

DR EMBL; AK122555; BAC65837.1; -.

DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0045285; C:ubiquinol-cytochrome-c reductase complex; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR GO; GO:0008121; F:ubiquinol-cytochrome-c reductase activity; IEA.

DR GO; GO:0006118; P:electron transport; IEA.

DR InterPro; IPR006209; EGF_like.

DR InterPro; IPR006210; IEGF.

DR InterPro; IPR002049; Laminin_EGF.

DR InterPro; IPR005805; Rieske.

DR Pfam; PF00008; EGF; 10.

DR PRINTS; PR00011; EGFLAMININ.

DR SMART; SM00181; EGF; 14.

DR SMART; SM00180; EGF_Lam; 14.

DR PROSITE; PS00022; EGF_1; 14.

DR PROSITE; PS01186; EGF_2; 14.

DR PROSITE; PS00200; RIESKE_2; 1.

FT NON_TER 1 1

SQ SEQUENCE 921 AA; 97316 MW; 60A34D9513A600F7 CRC64;

Query Match 43.7%; Score 1575; DB 11; Length 921;

Best Local Similarity 65.7%; Pred. No. 4.1e-142;

Matches 237; Conservative 35; Mismatches 89; Indels 0; Gaps 0;

```
Qy      225 CPPGKHGPQCEQRCPCQNGGVCHHVTGECSCPSGWMGTVCQGQPCPEGRFGKNCSQECQCH 284
      |||| || || ||||| |||:||||:| || || |||| | ||:||||:| ||
Db      1  CPPGSHGAHCELRCPCQNGGTCHHITGECACPPGWTGAVCAQPCPPGTFGQNCSDCPCH 60

Qy      285 NGGTCDAAATGQCHCSPGYTGERCQDECPVGTYGVLCAETCQCVNGGKCYHVSACGLCEAG 344
      :|| || |||||: || |:|:|:| ||:| ||:| ||:| ||:| ||:| ||:|
Db      61 HGGQCDHVTGQCHCTAGYMGDRCEECFPGTFGFLCSQRCDCHNGGQCSPATGACECEPG 120

Qy      345 FAGERCEARLCPEGLYGIKCDKRCPCCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGF 404
      : | |: |||||:| | || || ||||:| | |:| |||| ||||:| |:
Db      121 YKGPSCQERLCPEGLHGPCTLPCCDTEINTISCHPVTGACTCQPGWSGHYCNESCPAGY 180

Qy      405 YGEACQQICSCQNGADCDSTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVC 464
      || || |:||||| |:| ||||| | |: || |||| |||| | | | |
Db      181 YGNQCQLPCTCQNGADCHSITGSCTCAPGFMGEVCAVPCAAGTYGPNCSSVSCSNGGTC 240

Qy      465 SPVDGSCTCKAGWHGVDCSIRCPSGTWGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEK 524
      ||||| |:| |:|: ||||| || || || || |: ||: | || | |:
Db      241 SPVDGSCTCREGWQGLDCSLPCPSGTWGLNCNETCICANGAACSPFDGSCACTPGWLGD 300

Qy      525 CELPCQDGTGYNCAERCDCSHADGCHPTTGHCRCLPGWSGVHCDSDVCAEGRWGPNCPLP 584
      |||| |:|:| |:|:| ||||| | |||| || |:|: ||| | |||||:
Db      301 CELPCPDGTGFLNCSEHDCSHADGCDPVTGHCCCLAGWTGIRCDSTCPPGRWGPNCVS 360

Qy      585 C 585
      |
Db      361 C 361
```

RESULT 10

O88281

ID O88281 PRELIMINARY; PRT; 1574 AA.
AC O88281;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MEGF6.
GN MEGF6.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=98360089; PubMed=9693030;
RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple EGF-
RT like motifs by motif-trap screening.";
RL Genomics 51:27-34(1998).
DR EMBL; AB011532; BAA32462.1; -.
DR PIR; T13954; T13954.
DR HSSP; P00736; LAPQ.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.

DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00008; EGF; 20.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00179; EGF_CA; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 5.
 DR PROSITE; PS00022; EGF_1; 23.
 DR PROSITE; PS01186; EGF_2; 23.
 DR PROSITE; PS01187; EGF_CA; 5.
 KW EGF-like domain.
 SQ SEQUENCE 1574 AA; 165445 MW; 2B48533D8F77F6E7 CRC64;

Query Match 38.1%; Score 1372.5; DB 11; Length 1574;
 Best Local Similarity 42.2%; Pred. No. 1.7e-122;
 Matches 230; Conservative 60; Mismatches 199; Indels 56; Gaps 11;

Qy	89	RRKSQCCPGFYESGEMCVPHCADKCVH-GRCIAPNT--CQCEPGWGGTNCSSACDGDHWG	145
		: : : : :	
Db	816	RCQDTCAGWYGTG--CQIRCA--CANDGHC-DPTTGRCSAPGWTGLSCQRACDSGHWG	870
Qy	146	PHCTSRCQCKNG-ALCNPITGACHCAAGFRGWRCEDRCEQGTGNDCHQRCQCQNGATCD	204
		: : : : : :	
Db	871	PDCIHPNCNSAGHGNDAVSGLCCLCEAGYEGPRCEQSCRQGYGSPCEQKCRCEHGAACD	930
Qy	205	HVTGECRCPPGYTGAFCE-----LCPPGKHGPPQC	234
		: : : : : :	
Db	931	HVSGACTCPAGWRGSGFCEHACPAFFGLDCDSACNCSAGAPCDAVTGSCICPAGRWGPRC	990
Qy	235	EQRCP-----CQNGGVCHHVTGECSCPSGWMGTVCQGPCPEGRFGKNCSQEC	281
		: :	
Db	991	AQSCPPLTFGLNCSQICTCFNGASCDSVTGQCHCAPGWMGPTCLQACPPGLYGKNCQHSC	1050
Qy	282	QCHNGGTCAATGQCHCSPGYTGERCQDECPVGTGVLCAETCQCVNNGGKCYHVSGACLC	341
		: : : : :	
Db	1051	LCRNGGRCDPILGQCTCPEGWTGLACENECLPGHYAAGCQLNCSCLHGGICDRLTGHCCLC	1110
Qy	342	EAGFAGERCEARLCPEGLYGIKCDKRCPCHELENTSCHPMSGECACKPGWSGLYCNETCS	401
		: : : : : : : :	
Db	1111	PAGWTGDKCQSS-CVSGTFGVHCEEHCAC--RKGASCHHVTGACFCPPGWRGPHCEQACP	1167
Qy	402	PGFYGEACQQICSCQNGADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKND	461
		: : : : :	
Db	1168	RGWFGEACAQRCLCPTNASCHHVTGECRCPPGFTGLSCEQACQPGTFGKDCEHLCQCPGE	1227
Qy	462	A-VCSPVDGSCTCKAGWHGVDCSIRCPSGTWGFGCNLTQCCLNGGACNTLDGTCTCAPGW	520
		: : : :	
Db	1228	TWACDPASGVCTCAAGYHGTGCLQRCPSGRYGPGCEHICKCLNGGTCDPATGACYCPAGF	1287
Qy	521	RGEKCELPQDGTGYNLCAERCDCSHADGCHPTTGHCRCLPGWSGVHCDVCAEGRWGPN	580
		: : : : : :	
Db	1288	LGADCSLACPQGRFGPSCAHVCACRQGAACDPVSGACICSPGKTGVRCEHGCPQDRFGKG	1347
Qy	581	CSLPC	585
Db	1348	CELKC	1352

Q8ND91

Query Match 37.8%; Score 1362; DB 4; Length 626;
Best Local Similarity 63.1%; Pred. No. 6.8e-122;
Matches 205; Conservative 40; Mismatches 80; Indels 0; Gaps 0;

Qy	261	GTVCGQPCPEGRFGKNCSEQECQCHNGGTCDAAATGQCHCSPGYTGERCQDECPVGTYGVL	320
		: : : : : :	
Db	6	GAVCAQPCPPGTFGQNCSDQDCPCHHGGQCDHVTGQCHCTAGYMGDRCQECPFGSFGFQC	65
Qy	321	AETCQCVNGGKCYHVSAGACLCEAGFAGERCEARLCPEGLYGIKCDKRCPCHLNTHSCHP	380
		: : : : : :	
Db	66	SQRCDCHNGGQCSPTTGACECEPGYKGPRCQERLCPEGLHGPGLTLPCCDADNTISCHP	125
Qy	381	MSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCDSDVTGKCTCAPGFKGIDCS	440
		: : : : : : : :	
Db	126	VTGACTCQPGWSGHHCNESCPVGYGDDGCLPCTCQNGADCHSITGGCTCAPGFMGEVCA	185

QY	441	TPCPLGTYGINCSSRCCKNDVAVCSVDGSCTCKAGWHGVDCSIRCPSGTWFGCNLTQ	500
		: :: :	
Db	186	VSCAAGTYGPNCSSICSCNNGGTCSVDGSCTCKEGWQGLDCTLPCPSGTWGLNCNESCT	245
QY	501	CLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAERCDCHADGCHPTTGHCRL	560
		: : :: : : :	
Db	246	CANGAACSPIDGSCSTPGWLGDTCLEPDPGTFLNCSEHCDCHADGCDPVTGHCCCL	305
QY	561	PGWSGVHCDSVCAEGRWGPNCSLPC	585
		:: :	
Db	306	AGWTGIRCDSTCPPGRWGPNCSVSC	330

RESULT 12

Q8BX64

```
ID      Q8BX64      PRELIMINARY;      PRT;      299 AA.
AC      Q8BX64;
DT      01-MAR-2003 (TrEMBLrel. 23, Created)
DT      01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      MEGF11 protein.
GN      2410080H04RIK.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX      MEDLINE=22354683; PubMed=12466851;
RA      The FANTOM Consortium,
RA      the RIKEN Genome Exploration Research Group Phase I & II Team;
RT      "Analysis of the mouse transcriptome based on functional annotation of
RT      60,770 full-length cDNAs.";
RL      Nature 420:563-573(2002).
DR      EMBL; AK048840; BAC33471.1; -.
DR      MGD; MGI:1920951; 2410080H04Rik.
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:0045285; C:ubiquinol-cytochrome-c reductase complex; IEA.
DR      GO; GO:0005198; F:structural molecule activity; IEA.
DR      GO; GO:0008121; F:ubiquinol-cytochrome-c reductase activity; IEA.
DR      GO; GO:0006118; P:electron transport; IEA.
DR      InterPro; IPR006209; EGF_like.
DR      InterPro; IPR006210; IEGF.
DR      InterPro; IPR002049; Laminin_EGF.
DR      InterPro; IPR005805; Rieske.
DR      Pfam; PF00008; EGF; 5.
DR      PRINTS; PR00011; EGFLAMININ.
DR      SMART; SM00181; EGF; 5.
DR      SMART; SM00180; EGF_Lam; 3.
DR      PROSITE; PS00022; EGF_1; 4.
DR      PROSITE; PS01186; EGF_2; 4.
DR      PROSITE; PS00200; RIESKE_2; 1.
SQ      SEQUENCE      299 AA;  32479 MW;  B5F27B185AE13D1A CRC64;
```

Query Match 37.2%; Score 1341; DB 11; Length 299;
Best Local Similarity 70.2%; Pred. No. 3.3e-120;

Matches 205; Conservative 34; Mismatches 53; Indels 0; Gaps 0;

```

Qy      15 LLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSCDILNWFKCTRHRV 74
      || : | || |||||:||||| ||||| | |||||:
Db      8 LLVFLQLAALALNPEDPNVCSHWESYAVTVQESYAHFPDQIYYTRCADILNWFKCTRHRI 67

Qy      75 SYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGWGGTN 134
      ||:|||| | :||||:||||:|:|: || | :|:||||:|:| ||||| :
Db      68 SYKTAYRRGLRTMYRRRSQCCPGYYENGDFCIPLCTEECMHGRCVSPDTCHCEPGWGGPD 127

Qy     135 CSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGNDCHQR 194
      ||| || :|||||:||||:||||| || |||||: | ||:| |
Db     128 CSSGCDSEHWGPHCSNRCQCNQNGALCNPITGACVCAPGFRGWRCEELCAPGTHGKGCQLL 187

Qy     195 CQCQNGATCDHVTGECRCPPGYTGAFCELDLCPGKHGPQCEQRCPCQNGGVCHHVTGECS 254
      ||| :||:| ||| | |||| :|:|||| || || ||||| |||:||||:
Db     188 CQCHHGASCDPRTGECLCAPGYTGVYCEELCPPGSHGAHCELCRCPCQNGGTCHHITGECA 247

Qy     255 CPSPGWMGTVCQGPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSPGYTGGER 306
      || || | || |||| | ||:||||:| ||:| || |||||: || |:|
Db     248 CPPGWTGAVCAQPCPPGTFGQNCSDQCPCHHGGQCDHVTGQCHCTAGYMGDR 299

```

RESULT 13

Q9TVQ2

```

ID   Q9TVQ2          PRELIMINARY;          PRT; 1664 AA.
AC   Q9TVQ2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Y64G10A.7 protein.
GN   Y64G10A.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mortimore B.J.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Ainscough R.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL117206; CAB60454.1; -.
DR   EMBL; AL110498; CAB60454.1; JOINED.
DR   EMBL; AL110498; CAB57911.1; -.
DR   EMBL; AL117206; CAB57911.1; JOINED.
DR   HSSP; P00736; 1APQ.
DR   WormPep; Y64G10A.7; CE24549.

```

DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00008; EGF; 22.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00179; EGF_CA; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 4.
 DR PROSITE; PS00022; EGF_1; 22.
 DR PROSITE; PS01186; EGF_2; 24.
 DR PROSITE; PS01187; EGF_CA; 3.
 KW EGF-like domain.
 SQ SEQUENCE 1664 AA; 179279 MW; A69F093B4C705832 CRC64;

Query Match 37.2%; Score 1340; DB 5; Length 1664;
 Best Local Similarity 36.4%; Pred. No. 2.3e-119;
 Matches 237; Conservative 70; Mismatches 218; Indels 126; Gaps 13;

QY	16	LCHWI-----GTASPL-----NLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC	60
		: : :	:
Db	824	VCHHVTGTCTCLPGKTGPLCDQSCAPNTYGPN-CAH-----TC	860
QY	61	TDILNWFKCTRHRVSYRTAYRHGEKTMYRRKSQCCPGFYESGEMCVPHCAD-----	111
		: :	
Db	861	S-CVNGAKCDESDGS-----CHCTPGFY--GATCSEVCPTGRFGIDCMQ	901
QY	112	--KCVHGR-CIAPN-TCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGAC	167
		: : : : : :	
Db	902	LCKCQNGAICDTSNGSCECAPGWSGKKCDKACAPGTFGKDCSKKCDCADGMHCDPSDGEC	961
QY	168	HCAAGFRGWRCEDRCEQGTYGNDCHQRCQCQNGAT-----	202
		: : : :	
Db	962	ICPPGKKGHKCDETCDSGLEFAGCKGICSCQNGATCDSVTGSCECRPGWRGKKCDRPCPD	1021
QY	203	-----CDHVTGECRCPPGYTGAFCEDLCPPGKHGPQC	234
		:	
Db	1022	GRFGEGCNAICDCTTTNDTSMYNPFVARCDHVTGECRCPAGWTGPDCQTSCLGRHGEHC	1081
QY	235	EQRCPCQNGGVCHHVTGECSCPSGWMGTVCQPCPEGRFGKNCSQECQCHNGGTCDAAATG	294
		: : : : :	
Db	1082	RHSCQCSNGASCDRVTGFCDPCPSGFMGKNCESECEPLWGSNCMKHCLCMHGGECKENG	1141
QY	295	QCHCSPGYTGERCQDECPVGTYGVLCAETCQCVNGGKCYHVSGLCEAGFAGERCEARL	354
		: : : : : :	
Db	1142	DCECIDGWTGPSCEFLCPFGQFGRNCAQRCNCKNGASCDRKTGRCECLPGWSGEHCE-KS	1200
QY	355	CPEGLYGIKCDKRCPCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICS	414
		: : : : :	
Db	1201	CVSGHYGAKCEETCEC--ENGALCDPISGHCSQCPGWRGKKCNRPCLKGYFGRHCSQSCR	1258
QY	415	CQNGADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDVCSVPDGSCTCK	474
		: : : : : : :	
Db	1259	CANSKSCDHISGRQCQPKGYAGHSCTELCPDGTFGESCSQKCDCGENSMCDALSGKCFCK	1318
QY	475	AGWHGVDCSIRCPSGTWGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTY	534

Db 1319 PGHSGSDCKSGCVQGRFGPDCNQLCSCENGVCDSSTGSCVCPGPGYIGTKCEIACQSDRF 1378
 QY 535 GLNCAERCDCSHADGCHPTTGHCRCLPGWSGVHCDSVCAEGRWGPNCSLPC 585
 Db 1379 GPTCEKICNCENGGTCDRLTGQCRCLPGFTGMTCNQVCPEGRFGAGCKEKC 1429

RESULT 14

Q63404

```

ID      Q63404          PRELIMINARY;          PRT;    220 AA.
AC      Q63404;
DT      01-NOV-1996 (TrEMBLrel. 01, Created)
DT      01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      (Clone REM4) ORF (Fragment).
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Holzman; TISSUE=Brain;
RX      MEDLINE=96235155; PubMed=8642059;
RA      Asakura K., Pogulis R.J., Pease L.R., Rodriguez M.;
RT      "A monoclonal autoantibody which promotes central nervous system
RT      remyelination is highly polyreactive to multiple known and novel
RT      antigens.";
RL      J. Neuroimmunol. 65:11-19(1996).
DR      EMBL; L41686; AAB05844.1; -.
DR      HSSP; P01132; 1EGF.
DR      GO; GO:0005198; F:structural molecule activity; IEA.
DR      InterPro; IPR006209; EGF_like.
DR      InterPro; IPR009030; Grow_fac_recep.
DR      InterPro; IPR002049; Laminin_EGF.
DR      Pfam; PF00008; EGF; 3.
DR      PRINTS; PR00011; EGFLAMININ.
DR      SMART; SM00180; EGF_Lam; 2.
DR      PROSITE; PS00022; EGF_1; 5.
DR      PROSITE; PS01186; EGF_2; 5.
KW      EGF-like domain; Laminin EGF-like domain.
FT      NON_TER      1      1
FT      NON_TER      220    220
SQ      SEQUENCE      220 AA;  23231 MW;  3119D391EAF64372 CRC64;

```

Query Match 36.4%; Score 1309; DB 11; Length 220;
Best Local Similarity 95.4%; Pred. No. 2.8e-117;
Matches 209; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

[illegible]

RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
 RA Tupy J.L., Bergman C.M., Berman B.P., Carlson J.W., Celniker S.E.,
 RA Clamp M.E., Drysdale R.A., Emmert D., Frise E., de Grey A.D.N.J.,
 RA Harris N.L., Kronmiller B., Marshall B., Millburn G.H., Richter J.,
 RA Russo S., Searle S.M.J., Smith E., Shu S., Smutniak F.,
 RA Whitfield E.J., Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J.,
 RA Lewis S.E.;
 RT "Annotation of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AE003472; AAF47553.2; -.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00008; EGF; 7.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00181; EGF; 12.
 DR SMART; SM00180; EGF_Lam; 11.
 DR PROSITE; PS00022; EGF_1; 11.
 DR PROSITE; PS01186; EGF_2; 13.
 DR PROSITE; PS00290; IG_MHC; 1.
 SQ SEQUENCE 881 AA; 96380 MW; 52196D164F52F5C1 CRC64;

Query Match 35.8%; Score 1290.5; DB 5; Length 881;
 Best Local Similarity 35.5%; Pred. No. 6.6e-115;
 Matches 210; Conservative 56; Mismatches 169; Indels 157; Gaps 3;

Qy 151 RCQCKNGALCNPTGACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCQNGATCDHVTGEC 210
 :| | | | :| | :| | | | :| | | | :| :| :| :| :| :| | | | |
 Db 2 QCDCLNNAVCEPFSGDCECAKGYTGARCADICPEGFFGANCSEKCRCEGGKCHHVSSEC 61
 Qy 211 RCPPGYTGAFCEDLCPGKHGPQCEQRCPCQNGGVCHHVTGECSCPSGWMGTVCQGQPCPE 270
 :| | | :| | :| | | | | | | | | | | | | | | | | | | | | | |
 Db 62 QCAPGFTGPLCDMRCPDGKHGAQCQCQDPCQNDGKCQPETGACMCNPGWTGDCVANKCPV 121
 Qy 271 GRFGKNCSQECQCHNGGTCDAAATGQCHCSPGYTGERCQDECPVGTYGVLCAETCQCVNGG 330
 | :| | :| :| | | | | | | | | | | | | | :| | | | | | | | | |
 Db 122 GSYGPGCQESCECYKGAPCHHITGQCECPPGYRGERCFCDECQLNTYGFNCMTCDCA 181

Qy	331	KCYHVSGACLCEAGFAGERCEARLCPEGLYGIKCDKRCPCHELENTHSCH-----	379
		: : : : : : : : : : : : : :	
Db	182	MCDRANGTCICNPGWTGAKCAERICEANKYGLDCNRTCECDMEHTDLCHPETGNCQCSIG	241
Qy	380	-----PMSGECACKPGWSGLYCNETCSPGFYGE	407
		: : : : : : : : : : : : : :	
Db	242	WSSAQCTRPCTFLRYGPNCELTCNCKNGAKCSPVNGTCLCAPGWRGPTCEESCEPGTFGQ	301
Qy	408	ACQQICSCQNGADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSPV	467
		: : : : : : : : : : : : : :	
Db	302	DCALRCDQCNGAKCEPETGQCLCTAGWKNIKCDRPCDLNHFGQDCAKVCDCNNNAACNPQ	361
Qy	468	DGSCCTCKAGW-----	477
		: : : : :	
Db	362	NGSCTCAAGWTGERCERKCDTGKFGHDCAQKCQCDFNNSLACDATNGRCVCKQDWGVCRC	421
Qy	478	-----HGVDCSIR-----	485
		: : :	
Db	422	LNNSSCDPDSGNCICSAGWTGADCAEPCPPGFYGMCKERCPEILHGNKSCDHITGEILC	481
Qy	486	-----CPSGTWGFGCNLTGQCLNGGACNTLDGTCTCAPGWRGEKCELPQDGT	533
		: : : : : : : : : : : : : :	
Db	482	RTGYIGLTCEHPCPAGLYGPGCKLKCNCHEGGECNHVTGQCQCLPGWTGSNCNESCPTDT	541
Qy	534	YGLNCAERCDCSHADGCHPTTGHCRCCLPGWSGVHCDSVCAEGRWGPNCSLPC	585
		: : : : : : : : : : : : :	
Db	542	YGQGAQRCRCVHHKVCRKADGMCICETGWSGTRCDEVCPGPFYGEHCMNTC	593

Search completed: March 26, 2004, 16:11:12
Job time : 30.5191 secs

OM protein - protein search, using sw model

Run on: March 26, 2004, 15:58:50 ; Search time 9.84589 Seconds
(without alignments)
3099.072 Million cell updates/sec

Title: US-10-092-390-4
Perfect score: 3601
Sequence: 1 MVISLNSCLS FICLLI CHWI.....HCDSVCAEGRWGPNC SLPCY 586

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	843.5	23.4	830	1	SREC_HUMAN	Q14162 homo sapien
2	813	22.6	833	1	SRC2_MOUSE	P59222 mus musculu
3	808	22.4	870	1	SRC2_HUMAN	Q96gp6 homo sapien
4	719	20.0	2524	1	NOTC_XENLA	P21783 xenopus lae
5	717	19.9	2437	1	NTC1_BRARE	P46530 brachydanio
6	697	19.4	2321	1	NTC3_HUMAN	Q9um47 homo sapien
7	693	19.2	2531	1	NTC1_MOUSE	Q01705 mus musculu
8	685.5	19.0	2318	1	NTC3_MOUSE	Q61982 mus musculu
9	685.5	19.0	2471	1	NTC2_RAT	Q9qw30 rattus norv
10	682.5	19.0	2319	1	NTC3_RAT	Q9r172 rattus norv
11	682	18.9	4289	1	TENX_HUMAN	P22105 homo sapien
12	677	18.8	2470	1	NTC2_MOUSE	O35516 mus musculu
13	677	18.8	2703	1	NOTC_DROME	P07207 drosophila
14	675.5	18.8	2003	1	NTC4_HUMAN	Q99466 homo sapien
15	675	18.7	2531	1	NTC1_RAT	Q07008 rattus norv
16	667.5	18.5	1213	1	JAG3_BRARE	Q90y54 brachydanio
17	666.5	18.5	2471	1	NTC2_HUMAN	Q04721 homo sapien

18	664.5	18.5	1064	1	FBP1_STRPU	P10079	strongyloce
19	662	18.4	2556	1	NTC1_HUMAN	P46531	homo sapien
20	658	18.3	1964	1	NTC4_MOUSE	P31695	mus musculu
21	646	17.9	1808	1	TENA_CHICK	P10039	gallus gall
22	644.5	17.9	2201	1	TENA_HUMAN	P24821	homo sapien
23	638	17.7	3695	1	LMA5_HUMAN	O15230	homo sapien
24	633	17.6	1238	1	JAG2_HUMAN	Q9y219	homo sapien
25	631.5	17.5	2139	1	CRB_DROME	P10040	drosophila
26	631	17.5	1247	1	JAG2_MOUSE	Q9qye5	mus musculu
27	629	17.5	1218	1	JAG1_HUMAN	P78504	homo sapien
28	619.5	17.2	1202	1	JAG2_RAT	P97607	rattus norv
29	616	17.1	1219	1	JAG1_RAT	Q63722	rattus norv
30	615	17.1	1242	1	JAG1_BRARE	Q90y57	brachydanio
31	614	17.1	1218	1	JAG1_MOUSE	Q9qxx0	mus musculu
32	611	17.0	1746	1	TENA_PIG	Q29116	sus scrofa
33	593.5	16.5	3672	1	LML2_CAEEL	Q21313	caenorhabdi
34	587	16.3	1408	1	SERR_DROME	P18168	drosophila
35	586	16.3	3718	1	LMA5_MOUSE	Q61001	mus musculu
36	577.5	16.0	1801	1	LMB2_RAT	P15800	rattus norv
37	576.5	16.0	3712	1	LMA_DROME	Q00174	drosophila
38	567	15.7	1798	1	LMB2_HUMAN	P55268	homo sapien
39	564.5	15.7	1799	1	LMB2_MOUSE	Q61292	mus musculu
40	561	15.6	3106	1	LMA2_MOUSE	Q60675	mus musculu
41	560.5	15.6	1429	1	LI12_CAEEL	P14585	caenorhabdi
42	559	15.5	3110	1	LMA2_HUMAN	P24043	homo sapien
43	556	15.4	833	1	DL_DROME	P10041	drosophila
44	546.5	15.2	3084	1	LMA1_MOUSE	P19137	mus musculu
45	536	14.9	1295	1	GLP1_CAEEL	P13508	caenorhabdi

ALIGNMENTS

RESULT 1

SREC_HUMAN

ID SREC_HUMAN STANDARD; PRT; 830 AA.

AC Q14162; O43701;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Endothelial cells scavenger receptor precursor (Acetyl LDL receptor)

DE (Scavenger receptor class F member 1).

GN SCARF1 OR SREC OR KIAA0149.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Umbilical vein endothelial cells;

RX MEDLINE=98058897; PubMed=9395444;

RA Adachi H., Tsujimoto M., Arai H., Inoue K.;

RT "Expression cloning of a novel scavenger receptor from human

RT endothelial cells.";

RL J. Biol. Chem. 272:31217-31220(1997).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=22086180; PubMed=11978792;
 RA Adachi H., Tsujimoto M.;
 RT "Characterization of the human gene encoding the scavenger receptor
 RT expressed by endothelial cell and its regulation by a novel
 RT transcription factor, endothelial zinc finger protein-2.";
 RL J. Biol. Chem. 277:24014-24021(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bone marrow;
 RX MEDLINE=96127530; PubMed=8590280;
 RA Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.;
 RT "Prediction of the coding sequences of unidentified human genes. IV.
 RT The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by
 RT analysis of cDNA clones from human cell line KG-1.";
 RL DNA Res. 2:167-174(1995).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- FUNCTION: Mediates the binding and degradation of acetylated low
 CC density lipoprotein (Ac-LDL). Mediates heterophilic interactions,
 CC suggesting a function as adhesion protein (By similarity).
 CC -!- SUBUNIT: Heterophilic interaction with SREC2 via its extracellular
 CC domain. The heterophilic interaction is suppressed by the presence
 CC of ligand such as Ac-LDL (By similarity).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
 CC -!- TISSUE SPECIFICITY: Endothelial cells.
 CC -!- SIMILARITY: Contains 6 EGF-like domains.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; D86864; BAA24070.1; -.
 DR EMBL; AB052946; BAC02692.1; -.
 DR EMBL; D63483; BAA09770.1; -.
 DR EMBL; BC039735; AAH39735.1; -.
 DR HSSP; P01180; 2BN2.
 DR Genew; HGNC:16820; SCARF1.
 DR MIM; 607873; -.
 DR GO; GO:0016021; C:integral to membrane; IDA.
 DR GO; GO:0030169; F:low-density lipoprotein binding; IDA.
 DR GO; GO:0004888; F:transmembrane receptor activity; TAS.
 DR GO; GO:0045192; P:low-density lipoprotein catabolism; TAS.
 DR GO; GO:0006898; P:receptor mediated endocytosis; TAS.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR009030; Grow_fac_recep.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR002049; Laminin_EGF.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00181; EGF; 5.
 DR PROSITE; PS00022; EGF_1; 6.
 DR PROSITE; PS01186; EGF_2; 6.
 DR PROSITE; PS50026; EGF_3; 3.
 KW Cell adhesion; Receptor; Repeat; Signal; Transmembrane;
 KW EGF-like domain; Glycoprotein.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 830 ENDOTHELIAL CELLS SCAVENGER RECEPTOR.
 FT DOMAIN 20 421 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 422 442 POTENTIAL.
 FT DOMAIN 443 830 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 53 87 EGF-LIKE 1.
 FT DOMAIN 95 130 EGF-LIKE 2.
 FT DOMAIN 155 191 EGF-LIKE 3.
 FT DOMAIN 215 249 EGF-LIKE 4.
 FT DOMAIN 302 339 EGF-LIKE 5.
 FT DOMAIN 351 382 EGF-LIKE 6.
 FT DOMAIN 476 620 PRO/SER-RICH.
 FT DOMAIN 622 798 GLY-RICH.
 FT DOMAIN 431 438 POLY-LEU.
 FT DISULFID 57 69 POTENTIAL.
 FT DISULFID 63 75 POTENTIAL.
 FT DISULFID 77 86 POTENTIAL.
 FT DISULFID 99 111 POTENTIAL.
 FT DISULFID 105 118 POTENTIAL.
 FT DISULFID 120 129 POTENTIAL.
 FT DISULFID 159 172 POTENTIAL.
 FT DISULFID 165 179 POTENTIAL.
 FT DISULFID 181 190 POTENTIAL.
 FT DISULFID 219 230 POTENTIAL.
 FT DISULFID 225 237 POTENTIAL.
 FT DISULFID 239 248 POTENTIAL.
 FT DISULFID 306 319 POTENTIAL.
 FT DISULFID 313 326 POTENTIAL.
 FT DISULFID 329 338 POTENTIAL.
 FT DISULFID 355 363 POTENTIAL.
 FT DISULFID 358 370 POTENTIAL.
 FT DISULFID 372 381 POTENTIAL.
 FT CARBOHYD 289 289 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 393 393 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 662 - 662 R -> W (IN REF. 3).
 SQ SEQUENCE 830 AA; 87430 MW; F560D9E1AA64D779 CRC64;

Query Match 23.4%; Score 843.5; DB 1; Length 830;
 Best Local Similarity 36.2%; Pred. No. 1.4e-47;
 Matches 158; Conservative 51; Mismatches 164; Indels 63; Gaps 15;

Qy 93 QCCPGFYESGEMC-VPHC--ADKCVHGR-CIAPNTCQCEPGWGGTNCSSACDGDHWGPHC 148
 ||| |: : : | : | | | | : | | :|:|: | :|| | | :||| |
 Db 40 QCCAGWRQKDQECTIPICEGPDACQKDEVCKVPGLCRCKPGFFGAHCSSRCPGQYWGPDC 99
 Qy 149 TSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCQNGATCDHVTG 208
 | | | | | | | | | | | | | | : | | | |
 Db 100 RESCPCHPHGQCEPATGACQCQADRWGARCEFPACACGPHGR-----CDPATG 146
 Qy 209 ECRCPPGYTGAFCEDLCPGKHGPQCEQRCPCQNGGVCHHVTGECSCPSGWMGTVCQGPC 268
 | | ||: : | | :||| | | | | |
 Db 147 VCHCEPGWWSSTCRRPCQCNTAAARCEQ-----ATGACVCKPGW----- 185
 Qy 269 PEGRFGKNCSQECQCHNGGTCDAAATGQCHCSPGYTGERCQDECPVGTYGVLCAETCQCVN 328
 :|: || | || | | : :|:| | ||: | | : | :||
 Db 186 ----WGRRCSFRCNCH-GSPCEQDSGRCACRPGWWGPECQQ-----CECVR 227
 Qy 329 GGKCYHVSAGACLCEAGFAGERCEARLCPEGLYGIKCDKRCPCCHLENTHSCHPMSGEC-AC 387
 |:| | | | | | | | | | :|:| | : : | | :| :|
 Db 228 -GRCSAASGECTCPPGFRGARCELP-CPAGSHGVQCAHSCG-RCKHNEPCSPDTGSCESC 284
 Qy 388 KPGWSGLYCNETCSPGFYGEACQQIC-SCQNGADCDSVTGKC-TCAPGFKGIDCSTPCPL 445
 :|||:| | : | || :||:|:| | | :|:| | : | | | | |||
 Db 285 EPGWNGTQCQQPCLPPTFGESCEQQCPHCRHGEACEPDTGHCQRCQDGPWLGPRCEDPCPT 344
 Qy 446 GTYGINCSSRCGCKNDVAVCSFVDSCTCKAGWHGVDCSIRCPSGTWGFGCNLTQCCLNGG 505
 ||:| :| | | | | | | | | :| | : | | :| :| | |
 Db 345 GTFGEDCGSTCPTCVQGSCTVTGDCVCSAGYWGPPSCNASCPAGFHGNNCSVPCECPE-G 403
 Qy 506 ACNTLDGTCTCAPGWR 521
 |: : |:| | |
 Db 404 LCHPVSGSCQPGSGSR 419

RESULT 2

SRC2_MOUSE

ID SRC2_MOUSE STANDARD; PRT; 833 AA.
 AC P59222;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Scavenger receptor class F member 2 precursor (Scavenger receptor
 DE expressed by endothelial cells 2 protein) (SREC-II).
 GN SCARF2 OR SREC2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.

RC STRAIN=C57BL/6J;
 RX MEDLINE=22267235; PubMed=12154095;
 RA Ishii J., Adachi H., Aoki J., Koizumi H., Tomita S., Suzuki T.,
 RA Tsujimoto M., Inoue K., Arai H.;
 RT "SREC-II, a new member of the scavenger receptor type F family,
 RT trans-interacts with SREC-I through its extracellular domain.";
 RL J. Biol. Chem. 277:39696-39702(2002).
 CC -!- FUNCTION: Probable adhesion protein, which mediates homophilic and
 CC heterophilic interactions. In contrast to SCARF1, it poorly
 CC mediates the binding and degradation of acetylated low density
 CC lipoprotein (Ac-LDL).
 CC -!- SUBUNIT: Homophilic and heterophilic interaction via its
 CC extracellular domain. Interacts with SCARF1. The heterophilic
 CC interaction with SCARF1, which is stronger than the homophilic
 CC interaction with itself, is suppressed by the presence of SCARF1
 CC ligand such as Ac-LDL.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
 CC -!- SIMILARITY: Contains 7 EGF-like domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF522197; AAN45862.1; -.
 DR MGD; MGI:1858430; Scarf2.
 DR GO; GO:0005044; F:scavenger receptor activity; IDA.
 DR GO; GO:0007157; P:heterophilic cell adhesion; IDA.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR002049; Laminin_EGF.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00181; EGF; 8.
 DR SMART; SM00180; EGF_Lam; 6.
 DR PROSITE; PS00022; EGF_1; 7.
 DR PROSITE; PS01186; EGF_2; 4.
 DR PROSITE; PS50026; EGF_3; 3.
 KW Cell adhesion; Receptor; Repeat; Signal; Transmembrane;
 KW EGF-like domain; Glycoprotein.
 FT SIGNAL 1 33 POTENTIAL.
 FT CHAIN 34 833 SCAVENGER RECEPTOR CLASS F MEMBER 2.
 FT DOMAIN 34 433 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 434 454 POTENTIAL.
 FT DOMAIN 455 791 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 68 102 EGF-LIKE 1.
 FT DOMAIN 114 145 EGF-LIKE 2.
 FT DOMAIN 146 174 EGF-LIKE 3.
 FT DOMAIN 175 204 EGF-LIKE 4.
 FT DOMAIN 205 233 EGF-LIKE 5.
 FT DOMAIN 234 262 EGF-LIKE 6.
 FT DOMAIN 364 395 EGF-LIKE 7.
 FT DOMAIN 639 714 PRO-RICH.
 FT DISULFID 72 84 POTENTIAL.
 FT DISULFID 78 90 POTENTIAL.

FT	DISULFID	92	101	POTENTIAL.
FT	DISULFID	118	126	POTENTIAL.
FT	DISULFID	120	133	POTENTIAL.
FT	DISULFID	135	144	POTENTIAL.
FT	DISULFID	148	155	POTENTIAL.
FT	DISULFID	150	162	POTENTIAL.
FT	DISULFID	164	173	POTENTIAL.
FT	DISULFID	177	185	POTENTIAL.
FT	DISULFID	179	192	POTENTIAL.
FT	DISULFID	194	203	POTENTIAL.
FT	DISULFID	207	214	POTENTIAL.
FT	DISULFID	209	221	POTENTIAL.
FT	DISULFID	223	232	POTENTIAL.
FT	DISULFID	236	243	POTENTIAL.
FT	DISULFID	238	250	POTENTIAL.
FT	DISULFID	252	261	POTENTIAL.
FT	DISULFID	368	376	POTENTIAL.
FT	DISULFID	371	383	POTENTIAL.
FT	DISULFID	385	394	POTENTIAL.
FT	CARBOHYD	75	75	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	302	302	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	357	357	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	395	395	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	833 AA; 87871 MW; 51EADDEEAACAFF005 CRC64;		

Query Match 22.6%; Score 813; DB 1; Length 833;
 Best Local Similarity 32.6%; Pred. No. 1.3e-45;
 Matches 154; Conservative 46; Mismatches 161; Indels 112; Gaps 17;

Qy	94	CCPGFYESGEMC--VPHCA--DKCVHGR-CIAPNTCQCEPGWGGTNCSSACDGDHWGPHCT	149
		: : : : : :	
Db	56	CCAGWRQLGDECGLAVCEGNSTCSENEVCVRPGECRCRHGYFGANCDTKCPRQFWGPDCK	115
Qy	150	SRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCQNGATCDHVTGE	209
		: : :	
Db	116	ERCSCHPHGQCEDVTGQCTCHA--RRW-----GARCEHACQCQHG--TCHPRSGA	161
Qy	210	CRCPPGYTGAFCEDLCPGKHGPQCEQRCPCQNGGVCHHVTGECSCPSGWMGTVCGQPCP	269
		:	
Db	162	CRCEPGWWGA-----QCASACYCSATSRCDPQTGACLCHVGW-----	198
Qy	270	EGRFGKNCSQECQCHNGGTCDAAATGQCHCSPGYTGERCQDECPVGTYGVLCAETCQCVNG	329
		: : : : : : :	
Db	199	---WGRSCNNQCAC--NSSPCEQQSGRCQCR-----	224
Qy	330	GKCYHVSAGACLEAGFAGERCEARLCPEGLYGIKCDKRCPCHELENTHS--CHPMSGECACK	388
		: : : : : :	
Db	225	-----ERMFGARCDRYCQC-----SHGRCHPVDGTACAD	253
Qy	389	PGWSGLYCNETCSPGFYGEACQQIC--SCQNGADCDSVTGKC--TCAPGFKGIDCSTPCPLG	446
		: : : : :	
Db	254	PGYRGKYCREPCPAGFYGPGCRRRCGQCKGQQPCTVVEGRCLTCEPGWNGTKCDQPCATG	313
Qy	447	TYGINCSSRC--GCKNDAVCSFVDGSCT--CKAGWHGVDCSIRCPSGTWGFGCNLTQCCLNG	504
		: : : : : :	
Db	314	FYEGCGHRCPPCRDGHACNHVTGKCTHCNAGWIGDRCETKCSNGTYGEDCAFCVSDCGS	373

CC interaction with SCARF1, which is stronger than the homophilic
 CC interaction with itself, is suppressed by the presence of SCARF1
 CC ligand such as Ac-LDL (By similarity).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
 CC -!- TISSUE SPECIFICITY: Predominantly expressed in endothelial cells.
 CC Expressed in heart, placenta, lung, kidney, spleen, small
 CC intestine and ovary.
 CC -!- SIMILARITY: Contains 7 EGF-like domains.
 CC -!- CAUTION: Ref.2 sequences differ from that shown due to
 CC frameshifts in positions 750, 751 and 768.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF522196; AAN45861.1; -.
 DR EMBL; BC000584; AAH00584.1; ALT_FRAME.
 DR EMBL; BC009326; AAH09326.1; ALT_FRAME.
 DR Genew; HGNC:19869; SCARF2.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR002049; Laminin_EGF.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00181; EGF; 7.
 DR SMART; SM00180; EGF_Lam; 6.
 DR PROSITE; PS00022; EGF_1; 7.
 DR PROSITE; PS01186; EGF_2; 4.
 DR PROSITE; PS50026; EGF_3; 3.
 KW Cell adhesion; Receptor; Repeat; Signal; Transmembrane;
 KW EGF-like domain; Glycoprotein; Polymorphism.
 FT SIGNAL 1 43 POTENTIAL.
 FT CHAIN 44 870 SCAVENGER RECEPTOR CLASS F MEMBER 2.
 FT DOMAIN 44 441 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 442 462 POTENTIAL.
 FT DOMAIN 463 830 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 76 110 EGF-LIKE 1.
 FT DOMAIN 122 153 EGF-LIKE 2.
 FT DOMAIN 154 182 EGF-LIKE 3.
 FT DOMAIN 183 212 EGF-LIKE 4.
 FT DOMAIN 213 241 EGF-LIKE 5.
 FT DOMAIN 242 270 EGF-LIKE 6.
 FT DOMAIN 372 403 EGF-LIKE 7.
 FT DOMAIN 652 851 PRO-RICH.
 FT DISULFID 80 92 POTENTIAL.
 FT DISULFID 86 98 POTENTIAL.
 FT DISULFID 100 109 POTENTIAL.
 FT DISULFID 126 134 POTENTIAL.
 FT DISULFID 128 141 POTENTIAL.
 FT DISULFID 143 152 POTENTIAL.
 FT DISULFID 156 163 POTENTIAL.
 FT DISULFID 158 170 POTENTIAL.
 FT DISULFID 172 181 POTENTIAL.
 FT DISULFID 185 193 POTENTIAL.

FT	DISULFID	187	200	POTENTIAL.
FT	DISULFID	202	211	POTENTIAL.
FT	DISULFID	215	222	POTENTIAL.
FT	DISULFID	217	229	POTENTIAL.
FT	DISULFID	231	240	POTENTIAL.
FT	DISULFID	244	251	POTENTIAL.
FT	DISULFID	246	258	POTENTIAL.
FT	DISULFID	260	269	POTENTIAL.
FT	DISULFID	376	384	POTENTIAL.
FT	DISULFID	379	391	POTENTIAL.
FT	DISULFID	393	402	POTENTIAL.
FT	CARBOHYD	83	83	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	310	310	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	365	365	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	403	403	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARIANT	777	777	D -> E (in dbSNP:759611).
FT				/FTId=VAR_015148.
FT	VARIANT	778	778	V -> L (in dbSNP:759612).
FT				/FTId=VAR_015149.
FT	VARIANT	819	819	A -> G (in dbSNP:874100).
FT				/FTId=VAR_015150.
FT	VARIANT	837	837	A -> G (in dbSNP:874101).
FT				/FTId=VAR_015151.
FT	CONFLICT	474	478	MISSING (IN REF. 2).
FT	CONFLICT	626	641	ALYARVARREARPARA -> GTRPTTTWITHSTAAS (IN
FT				REF. 2; AAH00584).
SQ	SEQUENCE	870 AA;	92479 MW;	DCB735A50E6E9D1F CRC64;

Query Match 22.4%; Score 808; DB 1; Length 870;
 Best Local Similarity 33.1%; Pred. No. 2.8e-45;
 Matches 156; Conservative 42; Mismatches 164; Indels 110; Gaps 17;

Qy	94	CCPGFYESGEMC--VPHCA--DKCVHGR-CIAPNTCQCEPGWGGTNCSSACDGDHWGPHCT	149
		: : : : : : : : : : : :	
Db	64	CCAGWRQQGDECGIAVCEGNSTCSENEVCVRPGECCRCRHGYFGANCDTKCPRQFWGPDCK	123
Qy	150	SRCQCKNGALCNPIITGACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCQNGATCDHVTGE	209
		: : : : : : : : : : :	
Db	124	ELCSCHPHGQCEDVTGQCTCHA--RRW-----GARCEHACQCQHG-TCHPRSGA	169
Qy	210	CRCPPGYTGAFCEDLCPGKGHPQCEQRCPCQNGGVCHHVTGECSCPSGWMGTVCQGP	269
		: : : : : : : : : :	
Db	170	CRCEPGWWGA-----QCASACYCSATSRCDPQTGACLCHAGW-----	206
Qy	270	EGREFGKNCSQECQCHNGGTCDAAATGQCHCSPGYTGERCQDECPVGTYGVLCETCQCVNG	329
		: :: : : : : : :	
Db	207	---WGRSCNNQCAC-NSSPCEQQSGRCQCR-----ER-----TFGARCDRYCQCFRG	249
Qy	330	GKCYHVSAGACLCEAGFAGERCEARLCPEGLYGIKCDKRCPCCHLENTHSCHPMSGECACKP	389
		: :	
Db	250	-----RCHPVDGTACACEP	262
Qy	390	GWSGLYCNETCSPGFYGEACQQIC-SCQNGADCDSVTGKC-TCAPGFKGIDCSTPCPLGT	447
		: :: : : : : : :	
Db	263	GYRGKYCREPCPAGFYGLGCRRCRGQCKGQQPCTVAEGRCLTCEPGWNGTKCDQPCATGF	322
Qy	448	YGINCSSRC-GCKNDAVCS PVDGSCT-CKAGWHGVDCSIRCPSGTWGFGCNLTQCCLNGG	505

```

      || || || |:: | : | | | | || | | : | : || : | | |
Db      323 YGEGCSHRCPPCRDGHACNHVTGKCTRCNAGWIGDRKETKCSNGTYGEDCAFCVADCGSG 382

Qy      506 ACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAERCDCSHADGCHPTTGHC 557
      | : | | | : || | | : | | : || : | | | | | | | |
Db      383 HCDFQSGRCLCSPGVHGPHCNVTCPPGLHGADCAQACSC-HEDTCDPVTGAC 433

```

RESULT 4

NOTC_XENLA

```

ID      NOTC_XENLA          STANDARD;          PRT; 2524 AA.
AC      P21783;
DT      01-MAY-1991 (Rel. 18, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Neurogenic locus notch protein homolog precursor (XOTCH protein).
GN      XOTCH.
OS      Xenopus laevis (African clawed frog).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Amphibia; Batrachia; Anura; Mesobatrachia; Pipioidea; Pipidae;
OC      Xenopodinae; Xenopus.
OX      NCBI_TaxID=8355;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=90385285; PubMed=2402639;
RA      Coffman C., Harris W., Kintner C.;
RT      "Xotch, the Xenopus homolog of Drosophila notch.";
RL      Science 249:1438-1441(1990).
RN      [2]
RP      REVISIONS TO 1759-1782.
RA      Kintner C.;
RL      Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
CC      -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC      -!- DEVELOPMENTAL STAGE: Expressed almost uniformly in early embryos.
CC      -!- SIMILARITY: Belongs to the NOTCH family.
CC      -!- SIMILARITY: Contains 36 EGF-like domains.
CC      -!- SIMILARITY: Contains 3 Lin/Notch repeats.
CC      -!- SIMILARITY: Contains 6 ANK repeats.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M33874; AAB02039.1; -.
DR      HSSP; P00740; 1EDM.
DR      InterPro; IPR002110; ANK.
DR      InterPro; IPR000152; Asx_hydroxyl_S.
DR      InterPro; IPR000742; EGF_2.
DR      InterPro; IPR001881; EGF_Ca.
DR      InterPro; IPR001438; EGF_II.
DR      InterPro; IPR006209; EGF_like.
DR      InterPro; IPR002049; Laminin_EGF.
DR      InterPro; IPR008297; Notch.

```

DR InterPro; IPR000800; Notch_dom.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 36.
 DR Pfam; PF00066; notch; 3.
 DR PIRSF; PIRSF002279; Notch; 1.
 DR PRINTS; PR00010; EGFBLOOD.
 DR PRINTS; PR00011; EGFLAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF_CA; 24.
 DR SMART; SM00004; NL; 2.
 DR PROSITE; PS50297; ANK_REP_REGION; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 23.
 DR PROSITE; PS00022; EGF_1; 34.
 DR PROSITE; PS01186; EGF_2; 29.
 DR PROSITE; PS50026; EGF_3; 36.
 DR PROSITE; PS01187; EGF_CA; 21.
 KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Signal; Glycoprotein.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 2524 NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG.
 FT DOMAIN 20 1728 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1729 1750 POTENTIAL.
 FT DOMAIN 1751 2524 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 20 57 EGF-LIKE 1.
 FT DOMAIN 58 99 EGF-LIKE 2.
 FT DOMAIN 102 140 EGF-LIKE 3.
 FT DOMAIN 141 177 EGF-LIKE 4.
 FT DOMAIN 179 215 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 217 254 EGF-LIKE 6.
 FT DOMAIN 256 292 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 294 332 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 334 370 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 371 409 EGF-LIKE 10.
 FT DOMAIN 411 449 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 451 487 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 489 525 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 527 563 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 565 600 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 602 638 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 640 675 EGF-LIKE 17.
 FT DOMAIN 677 713 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 715 750 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 752 788 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 790 826 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 828 866 EGF-LIKE 22.
 FT DOMAIN 868 904 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 906 942 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 944 980 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 982 1018 EGF-LIKE 26.
 FT DOMAIN 1020 1056 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1058 1094 EGF-LIKE 28.
 FT DOMAIN 1096 1142 EGF-LIKE 29.
 FT DOMAIN 1144 1180 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1182 1218 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1220 1264 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).

FT	DOMAIN	1266	1304	EGF-LIKE 33.
FT	DOMAIN	1306	1346	EGF-LIKE 34.
FT	DOMAIN	1347	1383	EGF-LIKE 35.
FT	DOMAIN	1386	1424	EGF-LIKE 36.
FT	REPEAT	1441	1478	LIN/NOTCH 1.
FT	REPEAT	1479	1520	LIN/NOTCH 2.
FT	REPEAT	1521	1560	LIN/NOTCH 3.
FT	REPEAT	1876	1919	ANK 1.
FT	REPEAT	1924	1953	ANK 2.
FT	REPEAT	1957	1987	ANK 3.
FT	REPEAT	1991	2020	ANK 4.
FT	REPEAT	2024	2053	ANK 5.
FT	REPEAT	2057	2086	ANK 6.
FT	DISULFID	22	35	BY SIMILARITY.
FT	DISULFID	29	45	BY SIMILARITY.
FT	DISULFID	47	56	BY SIMILARITY.
FT	DISULFID	62	74	BY SIMILARITY.
FT	DISULFID	68	87	BY SIMILARITY.
FT	DISULFID	89	98	BY SIMILARITY.
FT	DISULFID	106	117	BY SIMILARITY.
FT	DISULFID	111	128	BY SIMILARITY.
FT	DISULFID	130	139	BY SIMILARITY.
FT	DISULFID	145	156	BY SIMILARITY.
FT	DISULFID	150	165	BY SIMILARITY.
FT	DISULFID	167	176	BY SIMILARITY.
FT	DISULFID	183	194	BY SIMILARITY.
FT	DISULFID	188	203	BY SIMILARITY.
FT	DISULFID	205	214	BY SIMILARITY.
FT	DISULFID	221	232	BY SIMILARITY.
FT	DISULFID	226	242	BY SIMILARITY.
FT	DISULFID	244	253	BY SIMILARITY.
FT	DISULFID	260	271	BY SIMILARITY.
FT	DISULFID	265	280	BY SIMILARITY.
FT	DISULFID	282	291	BY SIMILARITY.
FT	DISULFID	298	311	BY SIMILARITY.
FT	DISULFID	305	320	BY SIMILARITY.
FT	DISULFID	322	331	BY SIMILARITY.
FT	DISULFID	338	349	BY SIMILARITY.
FT	DISULFID	343	358	BY SIMILARITY.
FT	DISULFID	360	369	BY SIMILARITY.
FT	DISULFID	375	386	BY SIMILARITY.
FT	DISULFID	380	397	BY SIMILARITY.
FT	DISULFID	399	408	BY SIMILARITY.
FT	DISULFID	415	428	BY SIMILARITY.
FT	DISULFID	422	437	BY SIMILARITY.
FT	DISULFID	439	448	BY SIMILARITY.
FT	DISULFID	455	466	BY SIMILARITY.
FT	DISULFID	460	475	BY SIMILARITY.
FT	DISULFID	477	486	BY SIMILARITY.
FT	DISULFID	493	504	BY SIMILARITY.
FT	DISULFID	498	513	BY SIMILARITY.
FT	DISULFID	515	524	BY SIMILARITY.
FT	DISULFID	531	542	BY SIMILARITY.
FT	DISULFID	536	551	BY SIMILARITY.
FT	DISULFID	553	562	BY SIMILARITY.
FT	DISULFID	569	579	BY SIMILARITY.
FT	DISULFID	574	588	BY SIMILARITY.

FT	DISULFID	590	599	BY SIMILARITY.
FT	DISULFID	606	617	BY SIMILARITY.
FT	DISULFID	611	626	BY SIMILARITY.
FT	DISULFID	628	637	BY SIMILARITY.
FT	DISULFID	644	654	BY SIMILARITY.
FT	DISULFID	649	663	BY SIMILARITY.
FT	DISULFID	665	674	BY SIMILARITY.
FT	DISULFID	681	692	BY SIMILARITY.
FT	DISULFID	686	701	BY SIMILARITY.
FT	DISULFID	703	712	BY SIMILARITY.
FT	DISULFID	719	729	BY SIMILARITY.
FT	DISULFID	724	738	BY SIMILARITY.
FT	DISULFID	740	749	BY SIMILARITY.
FT	DISULFID	756	767	BY SIMILARITY.
FT	DISULFID	761	776	BY SIMILARITY.
FT	DISULFID	778	787	BY SIMILARITY.
FT	DISULFID	794	805	BY SIMILARITY.
FT	DISULFID	799	814	BY SIMILARITY.
FT	DISULFID	816	825	BY SIMILARITY.
FT	DISULFID	832	843	BY SIMILARITY.
FT	DISULFID	837	854	BY SIMILARITY.
FT	DISULFID	856	865	BY SIMILARITY.
FT	DISULFID	872	883	BY SIMILARITY.
FT	DISULFID	877	892	BY SIMILARITY.
FT	DISULFID	894	903	BY SIMILARITY.
FT	DISULFID	910	921	BY SIMILARITY.
FT	DISULFID	915	930	BY SIMILARITY.
FT	DISULFID	932	941	BY SIMILARITY.
FT	DISULFID	986	997	BY SIMILARITY.
FT	DISULFID	991	1006	BY SIMILARITY.
FT	DISULFID	1008	1017	BY SIMILARITY.
FT	DISULFID	1024	1035	BY SIMILARITY.
FT	DISULFID	1029	1044	BY SIMILARITY.
FT	DISULFID	1046	1055	BY SIMILARITY.
FT	DISULFID	1062	1073	BY SIMILARITY.
FT	DISULFID	1067	1082	BY SIMILARITY.
FT	DISULFID	1084	1093	BY SIMILARITY.
FT	DISULFID	1100	1121	BY SIMILARITY.
FT	DISULFID	1115	1130	BY SIMILARITY.

Query Match 20.0%; Score 719; DB 1; Length 2524;
 Best Local Similarity 25.9%; Pred. No. 3.6e-39;
 Matches 225; Conservative 60; Mismatches 221; Indels 364; Gaps 51;

QY	5	LNSCLSFICL-----LLCHWIGTASPLNLED-----PNVCSHW-----ESYS	41
		: : : : : :	
Db	603	INECLSKPCINGGQCTDRENGYICTCPKGT'TGVNCETKIDDCASNLCNDNGKCIDKIDGYE	662
QY	42	VTVQESYPHPFDQIYYT-----SCTDILNWFKCTRHRVSYRTAYRHGEKTMYRR	90
		: : :	
Db	663	CTCEPGYTGKLCNININECDSNPCRNGGTCKDQINGFTCV-----	702
QY	91	KSQCCPGFYESGEMC---VPHC-ADKCVHGRC---IAPNTCQCEPGWGGTNC---SSACD	140
		: : : : : :	
Db	703	----CPDGYHD-HMCLSEVNECNPNPCIHGACHDGVNGYKCDCEAGWSGSNCDINNNECE	757
QY	141	GDHWGPHCTSRCQCKNGALCNPI'TGA--CHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCQ	198

```

      :      | | | | : | | | | | :      | : |      |
Db      758 SN-----PCMNGGTCKDMTGAYICTCKAGFSGPNCQ-----TNINECSSN-PCL 800

QY      199 NGATC-DHVTG-ECRCPPGYTGAFCEDLCPGKHGPQCEQRC---PCQNGGVCHH----V 249
      | | | | | : | | | | | : | | | | |
Db      801 NHGTCIDDVAGYKCNCLPYTGAICEAVLAP-----CAGSPCKNGGRCKESEDFE 850

QY      250 TGECSGPSGWMGTVCQPCPEGRFGKNCSEQEC---QCHNGGTCDAAATG---QCHCSPGYTG 304
      | | | | | | : | | | | | : | | | | |
Db      851 TFSCECPPGWQQTCTC-----EIDMNECVNRPCRNGATCQNTNGSYKCNCKPGYTG 900

QY      305 ERCQ---DECPVGTYGVLCAETCQCVNNGKCYHVSGA--CLCEAGFAGERCEARL----- 354
      | : | : | : | | | | | | | | | : | : |
Db      901 RNCMDIDDC-----QPNPCHNGGSCSDGINMFFCNCBPAGFRGPKCEEDINECAS 950

QY      355 -----CPEGLYGIKCDKRCPCBLE---NTHSCHPMSG---ECAC 387
      | | | | : | | | : | : | | |
Db      951 NPCKNGANCTDCVNSYTCTCQPGFSGIHCESNTPDCTESSCFNGGTC--IDGINTFTCQC 1008

QY      388 KPGWSGLYC---NE-----TCSPGFYGEACQQI---CS---CQ 416
      | : | | | | | | | | : | | | : | | |
Db      1009 PPGFTGSYCQHDINECDSPCLNGGTCQDSYGTYKCTCPQGYTGLNQCQLVRWCDSSPCK 1068

QY      417 NGADCDSVTG--KCTCAPGFKGIDCSTP----- 442
      | | | : | | : | : | |
Db      1069 NGGKCWQTNNFYRCECKSGWTGVYCDVPSVSCEVAAKQQGVDIVHLCRNSGMCVDTGNTH 1128

QY      443 ---CPLGTYGINCSSR---CG---CKNDAVCSVPDG--SCTCKAGWHGVDCS----- 483
      | | | | : | | | : | | | | : | | |
Db      1129 FCRCQAGYTGSYCEEQVDECSNPNCQNGATCTDYLGGSCECVAGYHGVNCSEEINECLS 1188

QY      484 -----IRCPSTWGFGCNLT---C-----QCLNGGACNTLDG 512
      | | | | | : | | | | |
Db      1189 HPCQNGGTCIDLINTYKCSCLRGTQGVHCEINVDCTPFYDSFTLEPKCFNNGKCIDRVG 1248

QY      513 --TCTCAPGWRGEKCE-----LPCQD-GTYGLNCAE-----RCDC-----SH 546
      | | | : | : | | | | | | | : | | |
Db      1249 GYNICPPPGFVGECRCEGDVNECLSNPCDSRGTQ--NCIQLVNDYRCECRQGFTRRCESV 1306

QY      547 ADGC-----HPTTGH-CRCLPGWSGVHCD----- 569
      | | | : | | : | | | : |
Db      1307 VDGCKGMPCRNGGTCAVASNTERGFICKCPPPGFDGATCEYDSRTCSNLRQNGGTCISVL 1366

QY      570 ----SVCAEGRWGPNC-----SLPCY 586
      | : | | | | | | |
Db      1367 TSSKVCVSEGYTGATCQYPVISPCASHPCY 1396

```

RESULT 5

NTC1_BRARE

ID NTC1_BRARE STANDARD; PRT; 2437 AA.

AC P46530;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Neurogenic locus notch homolog protein 1 precursor.

GN NOTCH1A OR NOTCH.

OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=94128602; PubMed=8297791;
 RA Bierkamp C., Campos-Ortega J.A.;
 RT "A zebrafish homologue of the Drosophila neurogenic gene Notch and
 RT its pattern of transcription during early embryogenesis.";
 RL Mech. Dev. 43:87-100(1993).
 CC -!- FUNCTION: Implicated in cell fate specifications during
 CC embryo development. May be involved in the formation of the
 CC neural plate, notochord and brain vesicles.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- DEVELOPMENTAL STAGE: Expressed in all cells in pregastrulation
 CC stages. During gastrulation is differentially expressed,
 CC accumulating predominantly in the prechordal mesoderm and
 CC notochord. At the end of gastrulation, expressed along the
 CC anterior-posterior axis including the developing neural plate
 CC and differentiating mesoderm. Also present in the developing
 CC brain and head regions.
 CC -!- SIMILARITY: Belongs to the NOTCH family.
 CC -!- SIMILARITY: Contains 36 EGF-like domains.
 CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -!- SIMILARITY: Contains 6 ANK repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X69088; CAA48831.1; -.
 DR PIR; S42612; S42612.
 DR HSSP; P00740; 1EDM.
 DR ZFIN; ZDB-GENE-990415-173; notch1a.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR008297; Notch.
 DR InterPro; IPR000800; Notch_dom.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 36.
 DR Pfam; PF00066; notch; 3.
 DR PIRSF; PIRSF002279; Notch; 1.
 DR PRINTS; PR00010; EGFBLOOD.
 DR PRINTS; PR00011; EGFLAMININ.
 DR PRINTS; PR01452; NOTCH.

DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF_CA; 22.
 DR SMART; SM00004; NL; 3.
 DR PROSITE; PS50297; ANK_REP_REGION; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 23.
 DR PROSITE; PS00022; EGF_1; 34.
 DR PROSITE; PS01186; EGF_2; 28.
 DR PROSITE; PS50026; EGF_3; 36.
 DR PROSITE; PS01187; EGF_CA; 22.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Neurogenesis; Repeat; ANK repeat;
 KW EGF-like domain; Transmembrane; Glycoprotein; Signal.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 2437 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
 FT DOMAIN 21 1724 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1725 1747 POTENTIAL.
 FT DOMAIN 1748 2437 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 21 57 EGF-LIKE 1.
 FT DOMAIN 58 98 EGF-LIKE 2.
 FT DOMAIN 101 138 EGF-LIKE 3.
 FT DOMAIN 139 175 EGF-LIKE 4.
 FT DOMAIN 177 215 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 217 254 EGF-LIKE 6.
 FT DOMAIN 256 292 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 294 332 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 334 370 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 371 409 EGF-LIKE 10.
 FT DOMAIN 411 449 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 451 487 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 489 524 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 526 562 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 564 599 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 601 637 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 639 674 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 676 712 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 714 749 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 751 787 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 789 825 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 827 865 EGF-LIKE 22.
 FT DOMAIN 867 903 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 905 941 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 943 979 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 981 1017 EGF-LIKE 26.
 FT DOMAIN 1019 1055 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1057 1093 EGF-LIKE 28.
 FT DOMAIN 1095 1141 EGF-LIKE 29.
 FT DOMAIN 1143 1179 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1181 1217 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1219 1263 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1265 1303 EGF-LIKE 33.
 FT DOMAIN 1305 1344 EGF-LIKE 34.
 FT DOMAIN 1346 1382 EGF-LIKE 35.
 FT DOMAIN 1385 1423 EGF-LIKE 36.
 FT REPEAT 1446 1486 LIN/NOTCH 1.
 FT REPEAT 1487 1520 LIN/NOTCH 2.
 FT REPEAT 1521 1561 LIN/NOTCH 3.

FT	REPEAT	1867	1910	ANK 1.
FT	REPEAT	1915	1944	ANK 2.
FT	REPEAT	1948	1978	ANK 3.
FT	REPEAT	1982	2011	ANK 4.
FT	REPEAT	2015	2044	ANK 5.
FT	REPEAT	2048	2077	ANK 6.
FT	DOMAIN	2265	2276	POLY-GLN (OPA-REPEAT).
FT	SITE	1656	1657	CLEAVAGE (BY FURIN-LIKE PROTEASE) (BY
FT				SIMILARITY).
FT	DISULFID	25	35	BY SIMILARITY.
FT	DISULFID	29	45	BY SIMILARITY.
FT	DISULFID	47	56	BY SIMILARITY.
FT	DISULFID	62	73	BY SIMILARITY.
FT	DISULFID	67	86	BY SIMILARITY.
FT	DISULFID	88	97	BY SIMILARITY.
FT	DISULFID	105	116	BY SIMILARITY.
FT	DISULFID	110	126	BY SIMILARITY.
FT	DISULFID	128	137	BY SIMILARITY.
FT	DISULFID	143	154	BY SIMILARITY.
FT	DISULFID	148	163	BY SIMILARITY.
FT	DISULFID	165	174	BY SIMILARITY.
FT	DISULFID	181	194	BY SIMILARITY.
FT	DISULFID	188	203	BY SIMILARITY.
FT	DISULFID	205	214	BY SIMILARITY.
FT	DISULFID	221	232	BY SIMILARITY.
FT	DISULFID	226	242	BY SIMILARITY.
FT	DISULFID	244	253	BY SIMILARITY.
FT	DISULFID	260	271	BY SIMILARITY.
FT	DISULFID	265	280	BY SIMILARITY.
FT	DISULFID	282	291	BY SIMILARITY.
FT	DISULFID	298	311	BY SIMILARITY.
FT	DISULFID	305	320	BY SIMILARITY.
FT	DISULFID	322	331	BY SIMILARITY.
FT	DISULFID	338	349	BY SIMILARITY.
FT	DISULFID	343	358	BY SIMILARITY.
FT	DISULFID	360	369	BY SIMILARITY.
FT	DISULFID	375	386	BY SIMILARITY.
FT	DISULFID	380	397	BY SIMILARITY.
FT	DISULFID	399	408	BY SIMILARITY.
FT	DISULFID	415	428	BY SIMILARITY.
FT	DISULFID	422	437	BY SIMILARITY.
FT	DISULFID	439	448	BY SIMILARITY.
FT	DISULFID	455	466	BY SIMILARITY.
FT	DISULFID	460	475	BY SIMILARITY.
FT	DISULFID	477	486	BY SIMILARITY.
FT	DISULFID	493	503	BY SIMILARITY.
FT	DISULFID	498	512	BY SIMILARITY.
FT	DISULFID	514	523	BY SIMILARITY.
FT	DISULFID	530	541	BY SIMILARITY.
FT	DISULFID	535	550	BY SIMILARITY.
FT	DISULFID	552	561	BY SIMILARITY.
FT	DISULFID	568	578	BY SIMILARITY.
FT	DISULFID	573	587	BY SIMILARITY.
FT	DISULFID	589	598	BY SIMILARITY.
FT	DISULFID	605	616	BY SIMILARITY.
FT	DISULFID	610	625	BY SIMILARITY.
FT	DISULFID	627	636	BY SIMILARITY.

FT	DISULFID	643	653	BY SIMILARITY.
FT	DISULFID	648	662	BY SIMILARITY.
FT	DISULFID	664	673	BY SIMILARITY.
FT	DISULFID	680	691	BY SIMILARITY.
FT	DISULFID	685	700	BY SIMILARITY.
FT	DISULFID	702	711	BY SIMILARITY.
FT	DISULFID	718	728	BY SIMILARITY.
FT	DISULFID	723	737	BY SIMILARITY.
FT	DISULFID	739	748	BY SIMILARITY.
FT	DISULFID	755	766	BY SIMILARITY.
FT	DISULFID	760	775	BY SIMILARITY.
FT	DISULFID	777	786	BY SIMILARITY.
FT	DISULFID	793	804	BY SIMILARITY.
FT	DISULFID	798	813	BY SIMILARITY.
FT	DISULFID	815	824	BY SIMILARITY.
FT	DISULFID	831	842	BY SIMILARITY.
FT	DISULFID	836	853	BY SIMILARITY.
FT	DISULFID	855	864	BY SIMILARITY.
FT	DISULFID	871	882	BY SIMILARITY.
FT	DISULFID	876	891	BY SIMILARITY.
FT	DISULFID	893	902	BY SIMILARITY.
FT	DISULFID	909	920	BY SIMILARITY.

Query Match 19.9%; Score 717; DB 1; Length 2437;
 Best Local Similarity 25.7%; Pred. No. 4.7e-39;
 Matches 221; Conservative 60; Mismatches 208; Indels 370; Gaps 49;

Qy	4	SINSCLS-----FICLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESY	48
		:: : : ::	
Db	601	NINECLSQPCRNGGTCQDRENAYICTCPKGTGTVNCEINIDD---CKR-----	645
Qy	49	PHFPDQIYYTSCTDILNWFKCTRHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMC---	105
		: :: :	
Db	646	-KPCD---YGKCIDKINGYECV-----CEPGY---SGSMCNIN	676
Qy	106	VPHCA-----DKCVHGRC---IAPNT	123
		: :	
Db	677	IDDCALNPCHNGGTCIDGVNSFTCLCPDGFDRDATCLSQHNECSSNPCTHGSCLDQINSYR	736
Qy	124	CQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGA--CHCAAGFRGWRCEDR	181
		: : :	
Db	737	CVCEAGWMGRNCDININ-----ECLSN-PCVNGGTCKDMTSGYLCTCRAGFSGPNCQMN	789
Qy	182	CEQGTYGNDCHQRCQCQNGATC-DHVTG-ECRCPPGYTGAFCEDLCPPGKHGPQCEQRCP	239
		: : :	
Db	790	I-----NECASN-PCINQGSCIDDVAGFKCNCMLPYTGEVCENVLAP-----CSPR-P	835
Qy	240	CQNGGVCHH----VTGECSCPSGWMGTVCGQPCPEGRFGKNCSQEC---QCHNGGTCDAA	292
		: : : :	
Db	836	CKNGGVCRESEDFQSFSCNPAGWQGQTCEVDI-----NECVRNPCTNGGVCENL	885
Qy	293	TG--QCHCSPGYTGERCQ---DECPVGTYGVLCAETCQCVNGGKCY-HVSG-ACLCEAGF	345
		: :	
Db	886	RGGFQCRNPGFTGALCENDIDDC-----EPNPCSNGGVCQDRVNGFVCVCLAGF	935
Qy	346	AGERCEARL-----CPEGLYGIKCDKRCPCHLENTSHSHP-	380
		: :	

Db 936 RGERCAEDIDECVSAPCRNGGNCCTDCVNSYTCSCPAGFSGINCEINTPDCTES--SCFNG 993
 Qy 381 -----MSGECACKPGWSGLYC-----NE-----TCSPGFYGEA 408
 | | | | | : | | | | | : |
 Db 994 GTCVDGISSFSCVCLPGFTGNYCQHDVNECDSRPCQNGGSCQDGYGTYKCTCPHGYTGLN 1053
 Qy 409 CQQT---CS---CQNGADC--DSVTGKCTCAPGFKGIDCSTP----- 442
 | | : | | : | | : | | | | |
 Db 1054 CQSLVRWCDSSPCKNGGSCWQQGASFTCQCASGWTGIYCDVPSVSCEVAARQQGVSVAVL 1113
 Qy 443 -----CPLGTYGINCSSRCG-----CKNDAVCSPVDG--SCTCKAGW 477
 | | | | : | : | | : | | | | :
 Db 1114 CRHAGQCVDAGNTHLCRCQAGYTGSYCQEQVDECQPNPCQNGATCTDYLGGYSCECVPGY 1173
 Qy 478 HGVDCS-----IRCPSTWGFGCNL---TC----- 499
 | | : | | | | | : |
 Db 1174 HGMNCSKEINECLSQPCQNGGTCIDLVTYKCSCPRGTOGVHCEIDIDDCSPSVDPLTGE 1233
 Qy 500 -QCLNGGACNTLDG--TCTCAPGWRGEKCE-----LPCQ-DGTYGLNCAE-----RC 542
 : | | | | | : | | : | | | | : | |
 Db 1234 PRCFNGGRCVDRVGGYGCVCAPAGFVGERCEGDVNECLSDPCDPSGSY--NCVQLINDFRC 1291
 Qy 543 DCSHA-----DGCHPT-----TGH---CRCLPGWSGVHCD----- 569
 : | : | | | | | : | | : | | : |
 Db 1292 ECRTGYTGKRCETVFNGCKDTPCKNGGTCAVASNTKHGYICKCQPGYSGSSCEYDSQSCG 1351
 Qy 570 -----SVCAEGRWGPNC 581
 : | | | |
 Db 1352 SLRCRNGATCVSGHLSPRC 1370

RESULT 6

NTC3_HUMAN

ID NTC3_HUMAN STANDARD; PRT; 2321 AA.
 AC Q9UM47; Q9UEB3; Q9UPL3; Q9Y6L8;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 3 precursor (Notch 3).
 GN NOTCH3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97032728; PubMed=8878478;
 RA Joutel A., Corpechot C., Ducros A., Vahedi K., Chabriat H., Mouton P.,
 RA Alamowitch S., Domenga V., Cecillion M., Marechal E., Maciazek J.,
 RA Vayssiere C., Cruaud C., Cabanis E.-A., Ruchoux M.M., Weissenbach J.,
 RA Bach J.-F., Boussier M.-G., Tournier-Lasserre E.;
 RT "Notch3 mutations in CADASIL, a hereditary adult-onset condition
 RT causing stroke and dementia."
 RL Nature 383:707-710(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Gunel M., Artavanis-Tsakonas S.;

RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
 RA Burkhart-Schultz K., Gordon L., Kyle A., Ramirez M., Stilwagen S.,
 RA Phan H., Velasco N., Garnes J., Danganan L., Poundstone P.,
 RA Christensen M., Georgescu A., Avila J., Liu S., Attix C., Andreise T.,
 RA Trankheim M., Amico-Keller G., Coefield J., Duarte S., Lucas S.,
 RA Bruce R., Thomas P., Quan G., Kronmiller B., Arellano A.,
 RA Montgomery M., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.S.,
 RA Carrano A.V.;
 RT "Sequence analysis of an 1.5 Mb olfactory receptor (OLFR) cluster in
 RT 19p13.1.";
 RL Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
 RN [4]
 RP VARIANTS CADASIL TYR-49; CYS-71; CYS-90; CYS-110; CYS-133; CYS-141;
 RP ARG-146; CYS-153; CYS-169; CYS-171; CYS-182; ARG-185; SER-212;
 RP GLY-222; TYR-224; CYS-258; TYR-542; CYS-558; CYS-578; CYS-728;
 RP CYS-985; CYS-1006; CYS-1031; CYS-1231 AND ARG-1261, AND VARIANTS
 RP ARG-170; LEU-496; GLN-1133; MET-1183 AND ALA-2223.
 RX MEDLINE=98049753; PubMed=9388399;
 RA Joutel A., Vahedi K., Corpechot C., Troesch A., Chabriat H.,
 RA Vayssiere C., Cruaud C., Maciazek J., Weissenbach J., Bousser M.-G.,
 RA Bach J.-F., Tournier-Lasserre E.;
 RT "Strong clustering and stereotyped nature of Notch3 mutations in
 RT CADASIL patients.";
 RL Lancet 350:1511-1515(1997).
 RN [5]
 RP VARIANT CADASIL 114-GLY--PRO-120 DEL.
 RX MEDLINE=20264473; PubMed=10802807;
 RA Joutel A., Chabriat H., Vahedi K., Domenga V., Vayssiere C.,
 RA Ruchoux M.M., Lucas C., Leys D., Bousser M.-G., Tournier-Lasserre E.;
 RT "Splice site mutation causing a seven amino acid Notch3 in-frame
 RT deletion in CADASIL.";
 RL Neurology 54:1874-1875(2000).
 RN [6]
 RP IDENTIFICATION OF LIGANDS.
 RX MEDLINE=99180765; PubMed=10079256;
 RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
 RA Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
 RT "Human ligands of the Notch receptor.";
 RL Am. J. Pathol. 154:785-794(1999).
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Deltal to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity).
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -!- TISSUE SPECIFICITY: Ubiquitously expressed in fetal and adult
 CC tissues.
 CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form

CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-Golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane (By similarity).

CC -!- PTM: Phosphorylated (By similarity).

CC -!- DISEASE: Defects in NOTCH3 are associated with cerebral autosomal
CC dominant arteriopathy with subcortical infarcts and
CC leukoencephalopathy (CADASIL) [MIM:125310]. CADASIL causes a type
CC of stroke and dementia of which key features include recurrent
CC subcortical ischemic events and vascular dementia.

CC -!- SIMILARITY: Belongs to the NOTCH family.

CC -!- SIMILARITY: Contains 34 EGF-like domains.

CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.

CC -!- SIMILARITY: Contains 5 ANK repeats.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; U97669; AAB91371.1; -.

DR EMBL; AF058900; AAC14346.1; -.

DR EMBL; AF058881; AAC14346.1; JOINED.

DR EMBL; AF058882; AAC14346.1; JOINED.

DR EMBL; AF058883; AAC14346.1; JOINED.

DR EMBL; AF058884; AAC14346.1; JOINED.

DR EMBL; AF058885; AAC14346.1; JOINED.

DR EMBL; AF058886; AAC14346.1; JOINED.

DR EMBL; AF058887; AAC14346.1; JOINED.

DR EMBL; AF058888; AAC14346.1; JOINED.

DR EMBL; AF058889; AAC14346.1; JOINED.

DR EMBL; AF058890; AAC14346.1; JOINED.

DR EMBL; AF058891; AAC14346.1; JOINED.

DR EMBL; AF058892; AAC14346.1; JOINED.

DR EMBL; AF058893; AAC14346.1; JOINED.

DR EMBL; AF058894; AAC14346.1; JOINED.

DR EMBL; AF058895; AAC14346.1; JOINED.

DR EMBL; AF058896; AAC14346.1; JOINED.

DR EMBL; AF058897; AAC14346.1; JOINED.

DR EMBL; AF058898; AAC14346.1; JOINED.

DR EMBL; AF058899; AAC14346.1; JOINED.

DR EMBL; AC004257; AAC04897.1; -.

DR EMBL; AC004663; AAC15789.1; ALT_INIT.

DR PIR; S78549; S78549.

DR HSSP; P00740; 1EDM.

DR Genew; HGNC:7883; NOTCH3.

DR MIM; 600276; -.

DR MIM; 125310; -.

DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR008297; Notch.
 DR InterPro; IPR000800; Notch_dom.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 34.
 DR Pfam; PF00066; notch; 3.
 DR PIRSF; PIRSF002279; Notch; 1.
 DR PRINTS; PR00010; EGFBLOOD.
 DR PRINTS; PR00011; EGFLAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF_CA; 19.
 DR SMART; SM00004; NL; 3.
 DR PROSITE; PS50297; ANK_REP_REGION; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 18.
 DR PROSITE; PS00022; EGF_1; 33.
 DR PROSITE; PS01186; EGF_2; 25.
 DR PROSITE; PS50026; EGF_3; 34.
 DR PROSITE; PS01187; EGF_CA; 16.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
 KW Disease mutation.
 FT SIGNAL 1 39 POTENTIAL.
 FT CHAIN 40 2321 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 3.
 FT CHAIN 1629 2321 NOTCH EXTRACELLULAR TRUNCATION (BY
 FT SIMILARITY).
 FT CHAIN 1662 2321 NOTCH INTRACELLULAR DOMAIN (BY
 FT SIMILARITY).
 FT DOMAIN 40 1643 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1644 1664 POTENTIAL.
 FT DOMAIN 1665 2321 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 40 77 EGF-LIKE 1.
 FT DOMAIN 78 118 EGF-LIKE 2.
 FT DOMAIN 119 156 EGF-LIKE 3.
 FT DOMAIN 158 195 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 197 234 EGF-LIKE 5.
 FT DOMAIN 236 272 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 274 312 EGF-LIKE 7.
 FT DOMAIN 314 350 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 351 389 EGF-LIKE 9.
 FT DOMAIN 391 429 EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 431 467 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 469 505 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 507 543 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 545 580 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 582 618 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 620 655 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 657 693 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 695 730 EGF-LIKE 18.

FT	DOMAIN	734	770	EGF-LIKE 19.
FT	DOMAIN	771	808	EGF-LIKE 20.
FT	DOMAIN	810	847	EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	849	885	EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	887	922	EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).

Query Match 19.4%; Score 697; DB 1; Length 2321;
 Best Local Similarity 25.2%; Pred. No. 8.7e-38;
 Matches 226; Conservative 61; Mismatches 249; Indels 360; Gaps 51;

Qy	5	LNSCLS-----FICLLLCHWIGTASPLNLED-----PNVC-SHWESY	40
		: : : : : : : : :	
Db	432	VNECLSGPCRQATCLDRIGQFTCICMAGFTGTYCEVDIDECQSSPCVNGGVCKDRVNGF	491
Qy	41	SVTVQESYPHPFDQIYYTSC--TDILNWFKCTRHRVSYRTAYRHGEKTMYYRRKSQCCPGF	98
		: : : :	
Db	492	SCTCPSGFSGSTCQLDVDECASTPCRNGAKCVDQPDGY-----ECRCAEGF	537
Qy	99	YESGEMC---VPHCA-DKCVHGRC---IAPNTCQCEPGWGGTNCSSACDGDHWGPHCTSR	151
		: : : : :	
Db	538	--EGTLCDRNVDDCSPDPCHHGRCVDGIASFSCACAPGYTGTRCESQVD-----ECRSQ	589
Qy	152	CQCKNGALCNPI TG--ACHCAAGFRGWRCE---DRCEQG--TYG--NDCHQR--CQCQNG	200
		: : : : :	
Db	590	-PCRHHGKCLDLVDKYLRCRPSGTTGVNCEVNIDDCASNPTFGVCRDGINRYDCVCQPG	648
Qy	201	AT---CDHVTGECRCPPGYTGAFCE-----LCPPGKHGPQC---EQRC---PCQNGG	244
		: :	
Db	649	FTGPLCNVEINECASSPCGEGGSCVDGENGFRCICPPGSLPPLCLPPSHPCAHEPCSH-G	707
Qy	245	VCHHVTG--ECSCPSGWMGTVCQ-----PCPEGRFGK	275
		: : :	
Db	708	ICYDAPGGFRCVCEPGWSGPRCSQSLARDACESQPCRAGGTCSSDGMGFHCTCPPGVQGR	767
Qy	276	NCS--QEC---QCHNGGTCAATGQ---CHCSPGYTGERCQ---DEC---PVGTYGVLC	320
		: : : : : :	
Db	768	QCELLSPCTPNPCEHGGRCESAPGQLPVCSCPPQGWQGPCRQQDVDECAGPAPCGPHGI-C	826
Qy	321	AE-----TCQ-----CVNGGKCYHVS-----ACLCEAGFAGERCEA	352
		: :	
Db	827	TNLAGSFSTCHGGYTGPSCDQDINDCDPNPCLNGGSCQDGVGSFSCSCLPGFAGPRC-A	885
Qy	353	R-----LCPEGLYGIKCDKRCPCHELENTHSCHPM-----	381
		: :	
Db	886	RDVDECLSNPCPGTCTDHVASFTCTCPPGYGGFHCEQDLP-----DCSPSSCFNGGT	938
Qy	382	-----SGEACKPGWSGLYCNE-----TCSPGFYGEACQ	410
		: : : :	
Db	939	CVDGVNSFSLCRPGYTGAHCQHEADPCLSRPCLHGGVCSAAHPGFRCTCLESFTGPQCQ	998
Qy	411	QI---CS---CQNGADCDSVTGKCTCAPGFKGIDC---STP-----	442
		: :	
Db	999	TLVDWCSRQPCQNGGRCVQTGAYCLCPPGWSGRLCDIRSLPCREAAAQIGVRLEQLCQAG	1058
Qy	443	-----CPLGTYGINCSSRCG-----CKNDAVCSFVDGS---CTCKAGWHGVD	481
		: : : : :	
Db	1059	GQCVDEDSSHYCVCPEGRTGSHCEQEVDPCLAQPCQHGGTCRGYMGGYMCECLPGYNGDN	1118

Qy 482 CS-----IRCPSGTWGFGCNLT---C-----QCLN 503
 | | | | | : | :||
 Db 1119 CEDDVDECASQPCQHGGSCIDLVARVLCSCPPGTLGVLC EINEDDCGPGPPLDSGPRCLH 1178
 Qy 504 GGACNTLDG--TCTCAPGWRGEKCEL---PCQDGTGYNCAERCDCSHADGCHPTTG--- 555
 | | | | | : | :|| | :| | |
 Db 1179 NGTCVDLVGGFRCTCPPGYTGLRCEADINECRSGA-----CHAAHTRDCLQDPGGGF 1230
 Qy 556 HCRCLPGWSGVHCDSV-----CAEGRWGPNC 581
 | | :|| | :| ||: ||| |
 Db 1231 RCLCHAGFSGPRCQTVLSPCESQPCQHGGQCRPSPGPGGLTFTCHCAQPFWGPRC 1286

RESULT 7

NTC1_MOUSE

ID NTC1_MOUSE STANDARD; PRT; 2531 AA.
 AC Q01705; Q06007; Q61905; Q99JC2; Q9QW58; Q9R0X7;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Neurogenic locus notch homolog protein 1 precursor (Notch 1) (Motch A)
 DE (mT14) (p300).
 GN NOTCH1 OR MOTCH.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Embryo;
 RX MEDLINE=93194170; PubMed=8449489;
 RA Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A.,
 RA Copeland N.G., Gridley T.;
 RT "Cloning, analysis, and chromosomal localization of Notch-1, a mouse
 RT homolog of Drosophila Notch."
 RL Genomics 15:259-264(1993).
 RN [2]
 RP SEQUENCE OF 731-1899 FROM N.A. (ISOFORM 2), AND DEVELOPMENTAL STAGE.
 RC STRAIN=CD-1; TISSUE=Embryo;
 RX MEDLINE=93050801; PubMed=1426644;
 RA Reaume A.G., Conlon R.A., Zirngibl R., Yamaguchi T.P., Rossant J.;
 RT "Expression analysis of a Notch homologue in the mouse embryo."
 RL Dev. Biol. 154:377-387(1992).
 RN [3]
 RP SEQUENCE OF 1551-1647 FROM N.A. (ISOFORM 1), AND DEVELOPMENTAL STAGE.
 RC TISSUE=Embryo;
 RX MEDLINE=93048835; PubMed=1425352;
 RA Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,
 RA Greenspan R.J., McMahon A.P., Gridley T.;
 RT "Expression pattern of Motch, a mouse homolog of Drosophila Notch,
 RT suggests an important role in early postimplantation mouse
 RT development."
 RL Development 115:737-744(1992).
 RN [4]
 RP SEQUENCE OF 1161-1547 FROM N.A.
 RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;

RX MEDLINE=93178563; PubMed=8440332;
 RA Lardelli M., Lendahl U.;
 RT "Notch A and Notch B-two mouse Notch homologues coexpressed in a
 RT wide variety of tissues.";
 RL Exp. Cell Res. 204:364-372(1993).
 RN [5]
 RP SEQUENCE OF 1659-1673 FROM N.A.
 RX MEDLINE=99364499; PubMed=10437788;
 RA Lee J.S., Ishimoto A., Yanagawa S.I.;
 RT "Murine leukemia provirus-mediated activation of the Notch1 gene leads
 RT to induction of HES-1 in a mouse T lymphoma cell line, DL-3.";
 RL FEBS Lett. 455:276-280(1999).
 RN [6]
 RP SEQUENCE OF 1950-2201 FROM N.A.
 RX MEDLINE=98029496; PubMed=9384671;
 RA Messerle M., Follo M., Nehls M., Eggert H., Boehm T.;
 RT "Dynamic changes in gene expression during in vitro differentiation of
 RT mouse embryonic stem cells.";
 RL Cytokines Cell. Mol. Ther. 1:139-143(1995).
 RN [7]
 RP SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND
 RP MUTAGENESIS OF 1651-ARG--ARG-1654.
 RX MEDLINE=98318619; PubMed=9653148;
 RA Logeat F., Bessia C., Brou C., LeBail O., Jarriault S., Seidah N.G.,
 RA Israel A.;
 RT "The Notch1 receptor is cleaved constitutively by a furin-like
 RT convertase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).
 RN [8]
 RP PARTIAL SEQUENCE, AND POST-TRANSLATIONAL PROCESSING.
 RX MEDLINE=21523956; PubMed=11518718;
 RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
 RT "Murine notch homologs (N1-4) undergo presenilin-dependent
 RT proteolysis.";
 RL J. Biol. Chem. 276:40268-40273(2001).
 RN [9]
 RP POST-TRANSLATIONAL PROCESSING.
 RX MEDLINE=21374376; PubMed=11459941;
 RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
 RT "Conservation of the biochemical mechanisms of signal transduction
 RT among mammalian Notch family members.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
 RN [10]
 RP INTERACTION WITH DTX1 AND DTX2.
 RX MEDLINE=21123790; PubMed=11226752;
 RA Kishi N., Tang Z., Maeda Y., Hirai A., Mo R., Ito M., Suzuki S.,
 RA Nakao K., Kinoshita T., Kadesch T., Hui C.-C., Artavanis-Tsakonas S.,
 RA Okano H., Matsuno K.;
 RT "Murine homologs of deltex define a novel gene family involved in
 RT vertebrate Notch signaling and neurogenesis.";
 RL Int. J. Dev. Neurosci. 19:21-35(2001).
 CC --!- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Deltal to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and

apoptotic programs (By similarity). May play an essential role in
 postimplantation development, probably in some aspect of cell
 specification and/or differentiation. May be involved in mesoderm
 development, somite formation and neurogenesis. Involved in the
 maturation of both CD4+ and CD8+ cells in the thymus.
 -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 terminal fragment N(EC) which are probably linked by disulfide
 bonds. Interacts with DTX1 and DTX2.
 -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
 proteolytical processing NICD is translocated to the nucleus.
 -!- ALTERNATIVE PRODUCTS:
 Event=Alternative splicing; Named isoforms=2;
 Name=1;
 IsoId=Q01705-1; Sequence=Displayed;
 Name=2;
 IsoId=Q01705-2; Sequence=VSP_001402, VSP_001403, VSP_001404;
 Note=No experimental confirmation available;
 -!- TISSUE SPECIFICITY: Highly expressed in the brain, lung and
 thymus. Expressed at lower levels in the spleen, bone-marrow,
 spinal cord, eyes, mammary gland, liver, intestine, skeletal
 muscle, kidney and heart.
 -!- DEVELOPMENTAL STAGE: First detected in the mesoderm at 7.5 dpc By
 8.5 dpc highly expressed in presomitic mesoderm, mesenchyme and
 endothelial cells, while much lower levels are seen in the
 neuroepithelium. Between 9.5-10.5 dpc expressed at high levels in
 the neuroepithelium. At 13.5 dpc expressed in the surface
 ectoderm, eye and developing whisker follicles.
 -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 which is proteolytically cleaved by a furin-like convertase in the
 trans-Golgi network before it reaches the plasma membrane to yield
 an active, ligand-accessible form. Cleavage results in a C-
 terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 ligand binding, it is cleaved by TNF-alpha converting enzyme
 (TACE) to yield a membrane-associated intermediate fragment called
 notch extracellular truncation (NEXT). This fragment is then
 cleaved by presenilin dependent gamma-secretase to release a
 notch-derived peptide containing the intracellular domain (NICD)
 from the membrane.
 -!- PTM: Phosphorylated.
 -!- SIMILARITY: Belongs to the NOTCH family.
 -!- SIMILARITY: Contains 36 EGF-like domains.
 -!- SIMILARITY: Contains 3 Lin/Notch repeats.
 -!- SIMILARITY: Contains 5 ANK repeats.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL; Z11886; CAA77941.1; -.
 DR EMBL; L02613; AAK14898.1; -.
 DR EMBL; X68278; CAA48339.1; -.
 DR EMBL; AJ238029; CAB40733.1; -.
 DR EMBL; X82562; CAA57909.1; -.

DR PIR; A46019; A46019.
 DR PIR; B49175; B49175.
 DR HSSP; P00740; 1EDM.
 DR MGD; MGI:97363; Notchl.
 DR GO; GO:0005887; C:integral to plasma membrane; IC.
 DR GO; GO:0005515; F:protein binding; IPI.
 DR GO; GO:0030154; P:cell differentiation; IMP.
 DR GO; GO:0007386; P:compartment specification; IMP.
 DR GO; GO:0007219; P:N signaling pathway; IC.
 DR GO; GO:0045944; P:positive regulation of transcription from P. . .; IDA.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR008297; Notch.
 DR InterPro; IPR000800; Notch_dom.
 DR Pfam; PF00023; ank; 7.
 DR Pfam; PF00008; EGF; 35.
 DR Pfam; PF00066; notch; 3.
 DR PIRSF; PIRSF002279; Notch; 1.
 DR PRINTS; PR00010; EGF_BLOOD.
 DR PRINTS; PR00011; EGFLAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF_CA; 24.
 DR SMART; SM00004; NL; 2.
 DR PROSITE; PS50297; ANK_REP_REGION; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 2.
 DR PROSITE; PS00010; ASX_HYDROXYL; 22.
 DR PROSITE; PS00022; EGF_1; 34.
 DR PROSITE; PS01186; EGF_2; 27.
 DR PROSITE; PS50026; EGF_3; 36.
 DR PROSITE; PS01187; EGF_CA; 21.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation;
 KW Alternative splicing.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 2531 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
 FT CHAIN 1711 2531 NOTCH EXTRACELLULAR TRUNCATION.
 FT CHAIN 1744 2531 NOTCH INTRACELLULAR DOMAIN.
 FT DOMAIN 19 1725 EXTRACELLULAR (POTENTIAL).

Query Match 19.2%; Score 693; DB 1; Length 2531;
 Best Local Similarity 25.3%; Pred. No. 1.7e-37;
 Matches 217; Conservative 71; Mismatches 206; Indels 364; Gaps 51;

Qy 10 SFICLLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSCDILNWFKC 69
 |::|| | :||:| | :| | :| | ::::|
 Db 624 SYLCLCLKGTTGPNCEINLDD---CA-----SNPCDS---GTCLDKIDGYEC 664
 Qy 70 TRHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMC---VPHCA----- 110
 | ||: :| || : ||
 Db 665 A-----CEPGY--TGSMCNVNIDECAGSPCHNGGTCDGDIAG 699

Qy 111 -----DKCVHGRC---IAPNTCQCEPGWGGTNC---SSACDG 141
: | : | | : | | | | | : : | :
Db 700 FTRCRCPEGYHDP TCLSEVNECNSNPCIHGACRDGLNGYKCD CAPGWSGTNCDINNECES 759
Qy 142 DHWGPHTSRCQCKNGALCNPITG--ACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCN 199
: | | | : | | | | : | : | |
Db 760 N-----PCVNGGTCKDMTSGYVCTCREGFSGPNCQ-----TNINECASN-PCLN 802
Qy 200 GATC-DHVTG--ECRCPPGYTGAFCEDLCP PGKHGPQCEQRCPCQNGGVC----HHVTGEC 253
| | | | : | | | | : | | | | : : |
Db 803 QGTCIDDVAGYKCNCLPYTGATCEVVIAP-----C-ATSPCKNSGVCKESEDIYESFSC 855
Qy 254 SCPSGWMGTVC-----GQPCPEGR-----FGKNCS---QECQ--- 282
| : | | | | | | | : | | : :
Db 856 VCPTGWQQTCEVDINECVKSPCRHGASCQNTNGSYRCLCQAGYTGRNCESDIDDCRPNP 915
Qy 283 CHNGGTCDAA--TGQCHCSPGYTGERCQDE-----CPVGT 315
| | | | : | | | | : | | : : | | |
Db 916 CHNGGSCTDGIN TAFCDCLPGFQGA FCEEDINECASNPCQNGANCTDCVDSYTC TCFVGF 975
Qy 316 YGVLCAET-----CQCVNGGKCYHVS G----ACLCEAGFAGERCEARLCPEGLYGI-KC 364
| : | | | | | | | | : | : : |
Db 976 NGIHCE NNTPDCTESSCFNGGTC--VDGINSFTCLCPPGFTGSYCQ-----YDVNEC 1025
Qy 365 DKRCPCHE LTHSCHPM SG--ECACKPGWSGLYCNE-----TCS PGFYGEACQOI----- 412
| | | | | : | | : : | | : | | |
Db 1026 DSR-PCLHGGT--CQDSYGT YKCTCPQGYTG LNCQNLVRWCDSAPCKNGGRCWQTNTQYH 1082
Qy 413 CSCQN---GADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGIN-----CSSRCG----- 457
| : : | : | : | | : | | : : | : |
Db 1083 CECRSGWTGVNCDVLSVSCEVAAQKR GIDVTLLCQHGG LCVDEGDKHYCHCQAGYTGSYC 1142
Qy 458 -----CKNDAVCSPVDG--SCTCKAGWHGVDCS----- 483
| : | | : | | | | : | | : |
Db 1143 EDEVDECSNPCQNGATCTDYLG GFSCKCVAGYHGSNCSEEINECLSQPCQNGGT CIDLT 1202
Qy 484 ----IRCPSGTWGFGCNLT---C-----QCLNGGACNTLDG--TCTCAPGWRGE 523
| | | | | : | : | | | | | | : | |
Db 1203 NSYKCS CPRGTQGVHCEINVDDCHPPLDPASRS PKCFNNGTCVDQVGGYTCTCPPGFVGE 1262
Qy 524 KCE-----LPCQD-GTYGLNCAER-----CDC-----SHADGC----- 550
: | | | | | | : | : | : |
Db 1263 RCEGDVNECLSNPCDPRGTQ--NCVQRVND FHC ECRA GHTGRRCESVINGCRGKPKNGG 1320
Qy 551 -----HPTTGH-CRCLPGWSGVHCD S-----VCAEGRWGPNC----- 581
: | | | | : | | : : | | | |
Db 1321 VCAVASNTARGFICRC PAGFEGATCENDARTCGSLRCLNGGT CISGPRSP TCLCLGSFTG 1380
Qy 582 -----SLPCY 586
| | |
Db 1381 PECQFPASSPCVGSNPCY 1398

RESULT 8

NTC3_MOUSE

ID NTC3_MOUSE STANDARD; PRT; 2318 AA.

AC Q61982;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 3 precursor (Notch 3).
 GN NOTCH3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ICR X Swiss Webster;
 RX MEDLINE=95001556; PubMed=7918097;
 RA Lardelli M., Dalstrand J., Lendahl U.;
 RT "The novel Notch homologue mouse Notch 3 lacks specific epidermal
 RT growth factor-repeats and is expressed in proliferating
 RT neuroepithelium.";
 RL Mech. Dev. 46:123-136(1994).
 RN [2]
 RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1664.
 RX MEDLINE=21523956; PubMed=11518718;
 RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
 RT "Murine notch homologs (N1-4) undergo presenilin-dependent
 RT proteolysis.";
 RL J. Biol. Chem. 276:40268-40273(2001).
 RN [3]
 RP POST-TRANSLATIONAL PROCESSING.
 RX MEDLINE=21374376; PubMed=11459941;
 RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
 RT "Conservation of the biochemical mechanisms of signal transduction
 RT among mammalian Notch family members.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Deltal to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity). May play a role during CNS
 CC development.
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -!- TISSUE SPECIFICITY: Proliferating neuroepithelium.
 CC -!- DEVELOPMENTAL STAGE: CNS development.
 CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a

```

CC      notch-derived peptide containing the intracellular domain (NICD)
CC      from the membrane.
CC      -!- PTM: Phosphorylated.
CC      -!- SIMILARITY: Belongs to the NOTCH family.
CC      -!- SIMILARITY: Contains 34 EGF-like domains.
CC      -!- SIMILARITY: Contains 3 Lin/Notch repeats.
CC      -!- SIMILARITY: Contains 5 ANK repeats.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X74760; CAA52776.1; -.
DR      PIR; S45306; S45306.
DR      HSSP; P00740; 1EDM.
DR      MGD; MGI:99460; Notch3.
DR      GO; GO:0005887; C:integral to plasma membrane; IC.
DR      GO; GO:0005515; F:protein binding; IPI.
DR      GO; GO:0007219; P:N signaling pathway; IC.
DR      InterPro; IPR002110; ANK.
DR      InterPro; IPR000152; Asx_hydroxyl_S.
DR      InterPro; IPR000742; EGF_2.
DR      InterPro; IPR001881; EGF_Ca.
DR      InterPro; IPR001438; EGF_II.
DR      InterPro; IPR006209; EGF_like.
DR      InterPro; IPR002049; Laminin_EGF.
DR      InterPro; IPR008297; Notch.
DR      InterPro; IPR000800; Notch_dom.
DR      Pfam; PF00023; ank; 6.
DR      Pfam; PF00008; EGF; 33.
DR      Pfam; PF00066; notch; 3.
DR      PIRSF; PIRSF002279; Notch; 1.
DR      PRINTS; PR00010; EGF_BLOOD.
DR      PRINTS; PR00011; EGFLAMININ.
DR      PRINTS; PR01452; NOTCH.
DR      SMART; SM00248; ANK; 6.
DR      SMART; SM00179; EGF_CA; 19.
DR      SMART; SM00004; NL; 3.
DR      PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR      PROSITE; PS50088; ANK_REPEAT; 4.
DR      PROSITE; PS00010; ASX_HYDROXYL; 18.
DR      PROSITE; PS00022; EGF_1; 33.
DR      PROSITE; PS01186; EGF_2; 27.
DR      PROSITE; PS50026; EGF_3; 34.
DR      PROSITE; PS01187; EGF_CA; 16.
KW      Receptor; Transcription regulation; Activator; Differentiation;
KW      Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW      Transmembrane; Glycoprotein; Signal; Phosphorylation.
FT      SIGNAL          1      39      POTENTIAL.
FT      CHAIN           40     2318     NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 3.
FT      CHAIN          1630     2318     NOTCH EXTRACELLULAR TRUNCATION.
FT      CHAIN          1663     2318     NOTCH INTRACELLULAR DOMAIN.
FT      DOMAIN          40     1643     EXTRACELLULAR.

```


FT	TRANSMEM	1644	1664	POTENTIAL.
FT	DOMAIN	1665	2318	CYTOPLASMIC.
FT	DOMAIN	2242	2261	PEST.
FT	DOMAIN	40	78	EGF-LIKE 1.
FT	DOMAIN	79	119	EGF-LIKE 2.
FT	DOMAIN	120	157	EGF-LIKE 3.
FT	DOMAIN	159	196	EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	198	235	EGF-LIKE 5.
FT	DOMAIN	237	273	EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	275	313	EGF-LIKE 7.
FT	DOMAIN	315	351	EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	352	390	EGF-LIKE 9.
FT	DOMAIN	392	430	EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	432	468	EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	470	506	EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	508	544	EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	546	581	EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	583	619	EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	621	656	EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	658	694	EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	696	731	EGF-LIKE 18.
FT	DOMAIN	735	771	EGF-LIKE 19.
FT	DOMAIN	772	809	EGF-LIKE 20.
FT	DOMAIN	811	848	EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	850	886	EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	888	923	EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	925	961	EGF-LIKE 24.
FT	DOMAIN	963	999	EGF-LIKE 25.
FT	DOMAIN	1001	1035	EGF-LIKE 26.
FT	DOMAIN	1037	1083	EGF-LIKE 27.
FT	DOMAIN	1085	1121	EGF-LIKE 28.
FT	DOMAIN	1123	1159	EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1161	1204	EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1206	1245	EGF-LIKE 31.
FT	DOMAIN	1247	1288	EGF-LIKE 32.
FT	DOMAIN	1290	1326	EGF-LIKE 33.
FT	DOMAIN	1336	1374	EGF-LIKE 34.
FT	REPEAT	1388	1428	LIN/NOTCH 1.
FT	REPEAT	1429	1467	LIN/NOTCH 2.
FT	REPEAT	1468	1503	LIN/NOTCH 3.
FT	REPEAT	1839	1868	ANK 1.
FT	REPEAT	1872	1902	ANK 2.
FT	REPEAT	1906	1935	ANK 3.
FT	REPEAT	1939	1968	ANK 4.
FT	REPEAT	1972	2001	ANK 5.
FT	SITE	1572	1573	CLEAVAGE (BY FURIN-LIKE PROTEASE) (BY SIMILARITY).
FT				
FT	DISULFID	43	55	BY SIMILARITY.
FT	DISULFID	49	66	BY SIMILARITY.
FT	DISULFID	68	77	BY SIMILARITY.
FT	DISULFID	83	94	BY SIMILARITY.
FT	DISULFID	88	107	BY SIMILARITY.
FT	DISULFID	109	118	BY SIMILARITY.
FT	DISULFID	124	135	BY SIMILARITY.
FT	DISULFID	129	145	BY SIMILARITY.
FT	DISULFID	147	156	BY SIMILARITY.
FT	DISULFID	163	175	BY SIMILARITY.

FT	DISULFID	169	184	BY SIMILARITY.
FT	DISULFID	186	195	BY SIMILARITY.
FT	DISULFID	202	213	BY SIMILARITY.
FT	DISULFID	207	223	BY SIMILARITY.
FT	DISULFID	225	234	BY SIMILARITY.
FT	DISULFID	241	252	BY SIMILARITY.
FT	DISULFID	246	261	BY SIMILARITY.
FT	DISULFID	263	272	BY SIMILARITY.
FT	DISULFID	279	292	BY SIMILARITY.
FT	DISULFID	286	301	BY SIMILARITY.
FT	DISULFID	303	312	BY SIMILARITY.
FT	DISULFID	319	330	BY SIMILARITY.
FT	DISULFID	324	339	BY SIMILARITY.
FT	DISULFID	341	350	BY SIMILARITY.
FT	DISULFID	356	367	BY SIMILARITY.
FT	DISULFID	361	378	BY SIMILARITY.
FT	DISULFID	380	389	BY SIMILARITY.
FT	DISULFID	396	409	BY SIMILARITY.
FT	DISULFID	403	418	BY SIMILARITY.
FT	DISULFID	420	429	BY SIMILARITY.
FT	DISULFID	436	447	BY SIMILARITY.
FT	DISULFID	441	456	BY SIMILARITY.
FT	DISULFID	458	467	BY SIMILARITY.
FT	DISULFID	474	485	BY SIMILARITY.
FT	DISULFID	479	494	BY SIMILARITY.
FT	DISULFID	496	505	BY SIMILARITY.
FT	DISULFID	512	523	BY SIMILARITY.
FT	DISULFID	517	532	BY SIMILARITY.

Query Match 19.0%; Score 685.5; DB 1; Length 2318;
 Best Local Similarity 24.4%; Pred. No. 4.8e-37;
 Matches 216; Conservative 59; Mismatches 195; Indels 417; Gaps 48;

Qy	94	CCPGFYESGEMCVPHCADKCVHGRCIAPNT-----CQCEPGWGGTNCSSACDGDHW	144
		: : : :	
Db	225	CLPGF--EGQNCEVN-VDDCPGHRCLNGGTVCVDGVNTYNCQCPPEWTGQFCTEDVD----	277
Qy	145	GPHCTSRCQ-----CKNGALCNPITG--ACHCAAG-----FRGWRCE	179
		: :	
Db	278	-----ECQLQPNACHNGGTCTFNLLGGHSCVCVNGWTGESCSQNIDDCATAVCFHGATCH	331
Qy	180	DR-----CEQGTYGNDCH--QRC---QCQNGATCD--HVTGE--CRCPPGYTGAFCD	223
		: :	
Db	332	DRVASFYCACPMGKTGLLCHLDDACVSNPCHEDAICDTNPVSGRAICTCPPGFTGGACDQ	391
Qy	224	-----LCPPGKHGPQCE-----	235
		:	
Db	392	DVDECSIGANPCEHLGRCVNTQGSFLCQCGRGYTGPRCETDVNECLSGPCRNQATCLDRI	451
Qy	236	-----QRCPCQNGGVC--HHVTG--ECSCPSGWMGTVC-----	264
		: :	
Db	452	GQFTCICMAGFTGTYCEVDIDECQSSPCVNGGVCKDRVNGFSCTCPSGFSGSMCQLDVDE	511
Qy	265	-----GQP-----CPEGRFGKNCSQ---ECQ---CHNGGTCDATGQCHC	298
		: : :	
Db	512	CASTPCRNGAKCVDQPDGYECRCAEGFEGTLCERNVDDCSPDPCHHGRCVDGIASFSCAC	571

Qy 299 SPGYTGGERCQDE-----CPVGTYGVLC-----AETCQ 325
 :||||| ||: : || || || | : |
 Db 572 APGYTGIRCESQVDECRSQPCRYGGKCLDLVDKYLRCPPGTTGVNCEVNIDDCASNPT 631

Qy 326 ---CVNGGKCYHVSGACLCEAGFAGERCEARL-----CPEGLYGIKCDKRCP 369
 | :| | :||| | : | :| | | ||
 Db 632 FGVCRDGINRYD----CVCQPGFTGPLCNVEINECASSPCGEGGSCVDGENGFHC--LCP 685

Qy 370 -----CHLENTHS-CHPMSG--ECACKPGWSGLYCNE----- 398
 | :| || | | :||| | :|
 Db 686 PGSLLPPLCLPANHPCAHKPCSHGVCHDAPGGFRCVCEPGWSGPRCSQSLAPDACESQPCQ 745

Qy 399 -----TCSPGFYGEACQOI--CS---CQNGADCDSVTGK---CTCAPGFKG 436
 ||:||| | | : : | : | :| | : | :| ||:|
 Db 746 AGGTCTSDGIGFRCTCAPGFQGHQCEVLSPTPSLCEHGGHCESDPDRLTVCSCPPGWQG 805

Qy 437 -----IDCSTPCPLGTYGINCSS-----RCGCKNDAV-----CSP----- 466
 :|: | | :| | :| || | | |
 Db 806 PRCQQDVDECAGASPCGPHG--TCTNLPGNFRCICHRGYTGPFCDQDIDDCDPNPCLHGG 864

Qy 467 -VDG----SCTCKAGWHGVDC-----SIRCPSGTWGFGCNL 497
 || ||:| | :| | : | | | | :
 Db 865 CQDGVGSFSCSCLDGFAGPRCARDVDECLSSPCGPGTCTDHVASFTCACPPGYGGFHCEI 924

Qy 498 -----TCQCLNGGACNTLDG----TCTCAPGWRGEKC----- 525
 | ||| | :|| :| | ||: | |
 Db 925 DLPDCSPSSCFNGGTC--VDGVSSFSCLCRPGYTGTGHCQYEADPCFSRPCLHGGICNP 982

Qy 526 ---ELPCQDGTYGLNCAERCD-----CSHADGCHPTTGHCRCCLPGWSGVHCD----- 569
 | | :| | | | : | | | | | ||
 Db 983 PGFECTCREGFTGSQCQNPVDWCSQAPCQNGGRCVQTGAYCICPPGWSGRLCDIQSLPCT 1042

Qy 570 -----SVCAEGRWGPNCNL---PC 585
 || ||| | :| ||
 Db 1043 EAAAQMGVRLEQLCQEGGKCIDKGRSHYCVCPGRTGSHCEHEVDPC 1089

RESULT 9

NTC2_RAT

ID NTC2_RAT STANDARD; PRT; 2471 AA.
 AC Q9QW30;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 2 precursor (Notch 2).
 GN NOTCH2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=93202015; PubMed=1295745;
 RA Weinmaster G., Roberts V.J., Lemke G.;
 RT "Notch2: a second mammalian Notch gene."
 RL Development 116:931-941(1992).

RN [2]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=21331789; PubMed=11438922;
 RA Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
 RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
 RT functional roles for the Notch-DSL signaling system during brain
 RT development.";
 RL J. Comp. Neurol. 436:167-181(2001).
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs. May play an essential role in postimplantation
 CC development, probably in some aspect of cell specification and/or
 CC differentiation (By similarity).
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -!- TISSUE SPECIFICITY: Highly expressed in the spleen and choroid
 CC plexus in the brain. Expressed in postnatal central nervous system
 CC (CNS) germinal zones and, in early postnatal life, within numerous
 CC cells throughout the CNS. It is more highly localized to
 CC ventricular germinal zones. Also found in the heart, liver and
 CC kidney.
 CC -!- DEVELOPMENTAL STAGE: Expressed in the brain during E14 and E17.
 CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -!- PTM: Phosphorylated (By similarity).
 CC -!- SIMILARITY: Belongs to the NOTCH family.
 CC -!- SIMILARITY: Contains 35 EGF-like domains.
 CC -!- SIMILARITY: Contains 2 Lin/Notch repeats.
 CC -!- SIMILARITY: Contains 6 ANK repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M93661; AAK13558.1; -.
 DR PIR; A49128; A49128.
 DR HSSP; P00743; 1CCF.

DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR008297; Notch.
 DR InterPro; IPR000800; Notch_dom.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 35.
 DR Pfam; PF00066; notch; 2.
 DR PIRSF; PIRSF002279; Notch; 1.
 DR PRINTS; PR00010; EGF_BLOOD.
 DR PRINTS; PR00011; EGFLAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF_CA; 24.
 DR SMART; SM00004; NL; 2.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 22.
 DR PROSITE; PS00022; EGF_1; 34.
 DR PROSITE; PS01186; EGF_2; 26.
 DR PROSITE; PS50026; EGF_3; 35.
 DR PROSITE; PS01187; EGF_CA; 22.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 2471 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.
 FT CHAIN 1666 2471 NOTCH EXTRACELLULAR TRUNCATION (BY
 FT SIMILARITY).
 FT CHAIN 1697 2471 NOTCH INTRACELLULAR DOMAIN (BY
 FT SIMILARITY).
 FT DOMAIN 26 1677 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1678 1698 POTENTIAL.
 FT DOMAIN 1699 2471 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 26 63 EGF-LIKE 1.
 FT DOMAIN 64 102 EGF-LIKE 2.
 FT DOMAIN 105 143 EGF-LIKE 3.
 FT DOMAIN 144 180 EGF-LIKE 4.
 FT DOMAIN 182 219 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 221 258 EGF-LIKE 6.
 FT DOMAIN 260 296 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 298 336 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 338 374 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 375 413 EGF-LIKE 10.
 FT DOMAIN 415 454 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 456 492 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 494 530 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 532 568 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 570 605 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 607 643 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 645 680 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 682 718 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 720 755 EGF-LIKE 19.

FT	DOMAIN	757	793	EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	795	831	EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	833	871	EGF-LIKE 22.
FT	DOMAIN	873	909	EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	911	947	EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	949	985	EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	987	1023	EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1025	1061	EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1063	1099	EGF-LIKE 28.
FT	DOMAIN	1101	1147	EGF-LIKE 29.
FT	DOMAIN	1149	1185	EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1187	1223	EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1225	1262	EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1264	1302	EGF-LIKE 33.
FT	DOMAIN	1304	1343	EGF-LIKE 34.
FT	DOMAIN	1374	1412	EGF-LIKE 35.
FT	DOMAIN	1645	1648	POLY-ALA.
FT	DOMAIN	1994	1997	POLY-LEU.
FT	DOMAIN	2426	2429	POLY-SER.
FT	DOMAIN	2446	2451	POLY-GLY.
FT	REPEAT	1420	1456	LIN/NOTCH 1.
FT	REPEAT	1503	1535	LIN/NOTCH 2.
FT	REPEAT	1827	1871	ANK 1.
FT	REPEAT	1876	1905	ANK 2.
FT	REPEAT	1909	1939	ANK 3.
FT	REPEAT	1943	1972	ANK 4.
FT	REPEAT	1976	2005	ANK 5.
FT	REPEAT	2009	2038	ANK 6.
FT	DISULFID	28	41	BY SIMILARITY.
FT	DISULFID	35	51	BY SIMILARITY.
FT	DISULFID	53	62	BY SIMILARITY.
FT	DISULFID	68	79	BY SIMILARITY.
FT	DISULFID	73	90	BY SIMILARITY.
FT	DISULFID	92	101	BY SIMILARITY.
FT	DISULFID	109	121	BY SIMILARITY.
FT	DISULFID	115	131	BY SIMILARITY.
FT	DISULFID	133	142	BY SIMILARITY.
FT	DISULFID	148	159	BY SIMILARITY.
FT	DISULFID	153	168	BY SIMILARITY.
FT	DISULFID	170	179	BY SIMILARITY.
FT	DISULFID	186	198	BY SIMILARITY.
FT	DISULFID	192	207	BY SIMILARITY.
FT	DISULFID	209	218	BY SIMILARITY.
FT	DISULFID	225	236	BY SIMILARITY.
FT	DISULFID	230	246	BY SIMILARITY.
FT	DISULFID	248	257	BY SIMILARITY.
FT	DISULFID	264	275	BY SIMILARITY.
FT	DISULFID	269	284	BY SIMILARITY.
FT	DISULFID	286	295	BY SIMILARITY.
FT	DISULFID	302	315	BY SIMILARITY.
FT	DISULFID	309	324	BY SIMILARITY.
FT	DISULFID	326	335	BY SIMILARITY.
FT	DISULFID	342	353	BY SIMILARITY.
FT	DISULFID	347	362	BY SIMILARITY.
FT	DISULFID	364	373	BY SIMILARITY.
FT	DISULFID	379	390	BY SIMILARITY.
FT	DISULFID	384	401	BY SIMILARITY.

FT	DISULFID	403	412	BY SIMILARITY.
FT	DISULFID	419	433	BY SIMILARITY.
FT	DISULFID	427	442	BY SIMILARITY.
FT	DISULFID	444	453	BY SIMILARITY.
FT	DISULFID	460	471	BY SIMILARITY.
FT	DISULFID	465	480	BY SIMILARITY.
FT	DISULFID	482	491	BY SIMILARITY.
FT	DISULFID	498	509	BY SIMILARITY.
FT	DISULFID	503	518	BY SIMILARITY.
FT	DISULFID	520	529	BY SIMILARITY.
FT	DISULFID	536	547	BY SIMILARITY.

Query Match 19.0%; Score 685.5; DB 1; Length 2471;
 Best Local Similarity 24.8%; Pred. No. 5e-37;
 Matches 218; Conservative 83; Mismatches 240; Indels 339; Gaps 55;

Qy	3	ISLNSCLSFICL---LLCH-----WIGTASPLNLE--DPNVCSHW-----ES	39
		:: : : : :	
Db	531	IDIDDCSSTPCLNGAKCIDHPNGYECQCATGFTGTLCDENIDNCDPDPCHHGQCQDGIDS	590
Qy	40	YSVTVQESYPHPF--DQI--YYTS-----CTDIILNWFKCT-----RHRVSY----	76
		: : : : :	
Db	591	YTCICNPGYMGAIQSDQIDECYSSPCLNDGRCLDLVNGYQCNCQPGTSGLNCEINFDDCA	650
Qy	77	RTAYRHGE--KTMYYRRKSQCCPGFYESGEMC---VPHCADK-----	112
		: : :	
Db	651	SNPCLHGACVDGINRYSCVCSPGF--TGQRCNIDIDECASNPCRKDATCINDVNGFRMC	708
Qy	113	-----CVHGRC---IAPNTCQCEPGWGGTNCSSACDGDHWGPHCT	149
		: : : :	
Db	709	PEGPHHPSCYSQVNECLSSPCIHGNCTGGLSGYKCLCDAGWVGINCE--VDKN----ECL	762
Qy	150	SRCQCKNGALCNPITGA--CHCAAGFRGWRCE---DRC-----EQGTYGNDCH-QRCQC	197
		: : : : :	
Db	763	SN-PCQNGGTNNLVNGYRCTCKKGFKGYNQVNIDECASNPCLNQGTCLDDVSGYTCHC	821
Qy	198	Q---NGATCDHVTGECRCPPGYTGAFCE-----LCPGKGHPQCE---QRC---P	239
		: : :	
Db	822	MLPYTGKNCQTVLAPCSPNPCENAAVCKEAPNFESFTCLCAPGWQGRCTVDVDECVSKP	881
Qy	240	CQNGGVCHHVTGE--CSCPSGWMGTVCQPCPEGRFGKNCSQEC---QCHNGGTC--DAA	292
		: : : : : :	
Db	882	CMNNGICHNTQGSYMCECPPGFSMDCEEDI-----NDCLANPCQNGGSCVDKVN	931
Qy	293	TGQCHCSPGYTGERCQDE-----CPVGTYGVLG---AETC	324
		: : : : :	
Db	932	TFSCCLPGFVGDKCQTDMECLSEPCKNGGTCSYVNSYTCTCPAGFHGVHCENNIDEC	991
Qy	325	---QCVNGGKCYHVSG---ACLCEAGFAGERC-----EARLCPEGLYGIKC	364
		: : : :	
Db	992	TESSCFNGGTC--VDGINSFSLCPVGFTGPFCLHDINECSSNPCLNSTGTCVDGLGTYRC	1049
Qy	365	-----DKRC-----PCHLENTHSCHPMSGECACKPGWSGLYCNE-----	398
		: : :	
Db	1050	TCPLGYTGKNCQTLVNLCSPPCKNKGTCQAEKARPRCLCPPGWDGAYCDVLNVSKAAA	1109
Qy	399	-----TCSPGFYGEACQ---ICS---CQNGADCDSVTG	426

```

          | | : | | : : | : | : | | | |
Db      1110 LQKGVFVEHLCQHSGICINAGNTHHCQCPLGYTGSYCEEQLDECASNPCQHGATCSDFIG 1169
Qy      427 --KCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSFVDGSCCTCKAGWHGVDCSI 484
          :| | | : : | : | | : : | : | : |
Db      1170 GYRCECVPGYQGVNCE-----YEVDECQNQPCQNGGTCTIDLNVNHFKCS----- 1212
Qy      485 RCPSGTWGFGC--NL-TC----QCLNGGAC-NTLDG-TCTCAPGWRGEKCE-----L 527
          || || | | | : | | || | | : : | : | | : | : |
Db      1213 -CPPGTRGLLCEENIDDCAGAPHCLNGGQCVDRIGGYSCRCLPGFAGERCEGDINECLSN 1271
Qy      528 PC-QDGTYGLNCAE-----RCDCSHA-----DGCH-----PTTG 555
          || :| : | : | : | | | | | | | |
Db      1272 PCSSEGS--LDCIQLKNNYQCVCRSAFTGRHCETFLDVCPQKPCLNGGTCAVASNVDPDGF 1329
Qy      556 HCRCLPGWSGVHCDVCAEGRW-----GPNCSLP 584
          ||| || : || | | | : : | : | |
Db      1330 ICRCPPGFSGARCQSSCGQVKCRRGEQCVHTASGPHCFCP 1369

```

RESULT 10

NTC3_RAT

```

ID      NTC3_RAT          STANDARD;          PRT; 2319 AA.
AC      Q9R172;
DT      28-FEB-2003 (Rel. 41, Created)
DT      28-FEB-2003 (Rel. 41, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Neurogenic locus notch homolog protein 3 precursor (Notch 3).
GN      NOTCH3.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Haritunians T., Boulter J., Weinmaster G., Schanen N.C.;
RT      "Rattus norvegicus mRNA for Notch 3.";
RL      Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
RN      [2]
RP      FUNCTION.
RX      MEDLINE=21094508; PubMed=11182080;
RA      Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,
RA      Honjo T.;
RT      "Notch1 and Notch3 instructively restrict bFGF-responsive multipotent
RT      neural progenitor cells to an astroglial fate.";
RL      Neuron 29:45-55(2001).
RN      [3]
RP      TISSUE SPECIFICITY.
RX      MEDLINE=21331789; PubMed=11438922;
RA      Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
RT      "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
RT      functional roles for the Notch-DSL signaling system during brain
RT      development.";
RL      J. Comp. Neurol. 436:167-181(2001).
CC      -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC      Jagged1, Jagged2 and Deltal to regulate cell-fate determination.
CC      Upon ligand activation through the released notch intracellular

```


CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity). Acts instructively to control
 CC the cell fate determination of CNS multipotent progenitor cells,
 CC resulting in astroglial induction and neuron/oligodendrocyte
 CC suppression.
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -!- TISSUE SPECIFICITY: Expressed in postnatal central nervous system
 CC (CNS) germinal zones and, in early postnatal life, within
 CC numerous cells throughout the CNS. It is more highly localized
 CC to ventricular germinal zones.
 CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -!- PTM: Phosphorylated (By similarity).
 CC -!- SIMILARITY: Belongs to the NOTCH family.
 CC -!- SIMILARITY: Contains 34 EGF-like domains.
 CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -!- SIMILARITY: Contains 5 ANK repeats.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AF164486; AAD46653.2; -.
 DR HSSP; P00740; 1EDM.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR008297; Notch.
 DR InterPro; IPR000800; Notch_dom.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 33.
 DR Pfam; PF00066; notch; 3.
 DR PIRSF; PIRSF002279; Notch; 1.
 DR PRINTS; PR00010; EGF_BLOOD.

DR PRINTS; PR00011; EGFLAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF_CA; 20.
 DR SMART; SM00004; NL; 3.
 DR PROSITE; PS50297; ANK_REP_REGION; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 18.
 DR PROSITE; PS00022; EGF_1; 33.
 DR PROSITE; PS01186; EGF_2; 26.
 DR PROSITE; PS50026; EGF_3; 34.
 DR PROSITE; PS01187; EGF_CA; 16.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation.
 FT SIGNAL 1 40 POTENTIAL.
 FT CHAIN 41 2319 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 3.
 FT CHAIN 1631 2319 NOTCH EXTRACELLULAR TRUNCATION (BY
 FT SIMILARITY).
 FT CHAIN 1664 2319 NOTCH INTRACELLULAR DOMAIN (BY
 FT SIMILARITY).
 FT DOMAIN 41 1645 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1646 1666 POTENTIAL.
 FT DOMAIN 1667 2319 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 41 79 EGF-LIKE 1.
 FT DOMAIN 80 120 EGF-LIKE 2.
 FT DOMAIN 121 158 EGF-LIKE 3.
 FT DOMAIN 160 197 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 199 236 EGF-LIKE 5.
 FT DOMAIN 238 274 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 276 314 EGF-LIKE 7.
 FT DOMAIN 316 352 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 353 391 EGF-LIKE 9.
 FT DOMAIN 393 431 EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 433 469 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 471 507 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 509 545 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 547 582 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 584 620 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 622 657 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 659 695 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 697 732 EGF-LIKE 18.
 FT DOMAIN 736 772 EGF-LIKE 19.
 FT DOMAIN 773 810 EGF-LIKE 20.
 FT DOMAIN 812 849 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 851 887 EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 889 924 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 926 962 EGF-LIKE 24.
 FT DOMAIN 964 1000 EGF-LIKE 25.
 FT DOMAIN 1002 1036 EGF-LIKE 26.
 FT DOMAIN 1038 1084 EGF-LIKE 27.
 FT DOMAIN 1086 1122 EGF-LIKE 28.
 FT DOMAIN 1124 1160 EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1162 1205 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1207 1246 EGF-LIKE 31.
 FT DOMAIN 1248 1289 EGF-LIKE 32.
 FT DOMAIN 1291 1327 EGF-LIKE 33.

FT	DOMAIN	1337	1375	EGF-LIKE 34.
FT	REPEAT	1384	1420	LIN/NOTCH 1.
FT	REPEAT	1425	1461	LIN/NOTCH 2.
FT	REPEAT	1469	1503	LIN/NOTCH 3.
FT	REPEAT	1840	1869	ANK 1.
FT	REPEAT	1873	1903	ANK 2.
FT	REPEAT	1907	1936	ANK 3.
FT	REPEAT	1940	1969	ANK 4.
FT	REPEAT	1973	2002	ANK 5.
FT	SITE	1573	1574	CLEAVAGE (BY FURIN-LIKE PROTEASE) (BY
FT				SIMILARITY).
FT	DISULFID	44	56	BY SIMILARITY.
FT	DISULFID	50	67	BY SIMILARITY.
FT	DISULFID	69	78	BY SIMILARITY.
FT	DISULFID	84	95	BY SIMILARITY.
FT	DISULFID	89	108	BY SIMILARITY.
FT	DISULFID	110	119	BY SIMILARITY.
FT	DISULFID	125	136	BY SIMILARITY.
FT	DISULFID	130	146	BY SIMILARITY.
FT	DISULFID	148	157	BY SIMILARITY.
FT	DISULFID	164	176	BY SIMILARITY.
FT	DISULFID	170	185	BY SIMILARITY.
FT	DISULFID	187	196	BY SIMILARITY.
FT	DISULFID	203	214	BY SIMILARITY.
FT	DISULFID	208	224	BY SIMILARITY.
FT	DISULFID	226	235	BY SIMILARITY.
FT	DISULFID	242	253	BY SIMILARITY.
FT	DISULFID	247	262	BY SIMILARITY.
FT	DISULFID	264	273	BY SIMILARITY.
FT	DISULFID	280	293	BY SIMILARITY.
FT	DISULFID	287	302	BY SIMILARITY.
FT	DISULFID	304	313	BY SIMILARITY.
FT	DISULFID	320	331	BY SIMILARITY.
FT	DISULFID	325	340	BY SIMILARITY.
FT	DISULFID	342	351	BY SIMILARITY.
FT	DISULFID	357	368	BY SIMILARITY.
FT	DISULFID	362	379	BY SIMILARITY.
FT	DISULFID	381	390	BY SIMILARITY.
FT	DISULFID	397	410	BY SIMILARITY.
FT	DISULFID	404	419	BY SIMILARITY.
FT	DISULFID	421	430	BY SIMILARITY.
FT	DISULFID	437	448	BY SIMILARITY.
FT	DISULFID	442	457	BY SIMILARITY.
FT	DISULFID	459	468	BY SIMILARITY.
FT	DISULFID	475	486	BY SIMILARITY.
FT	DISULFID	480	495	BY SIMILARITY.
FT	DISULFID	497	506	BY SIMILARITY.
FT	DISULFID	513	524	BY SIMILARITY.
FT	DISULFID	518	533	BY SIMILARITY.
FT	DISULFID	535	544	BY SIMILARITY.
FT	DISULFID	551	561	BY SIMILARITY.

Query Match 19.0%; Score 682.5; DB 1; Length 2319;
 Best Local Similarity 24.7%; Pred. No. 7.5e-37;
 Matches 219; Conservative 56; Mismatches 196; Indels 415; Gaps 49;

QY 94 CCPGFYESGEMCVPHCADKCVHGRCIAPNT-----CQCEPGWGGTNCSSACDGDHW 144

Db	226	CLPGF--EGQNCEVN-VDDCPGHRCLNGGTCVDGVNTYNCQCPPEWTGQFCTEDVD----	278
Qy	145	GPHCTSRCQ-----CKNGALCNPTIG--ACHCAAG-----FRGWRCE	179
Db	279	-----ECQLQPNACHNGGTCFNLLGGHSCVCVNGWTGESCSQNIDDCATAVCFHGATCH	332
Qy	180	DR-----CEQGTYGNDCH--QRC---QCQNGATCD--HVTGE--CRCPPGYTGAFCD	223
Db	333	DRVASFYCACPMGKTGLLCHLDDACVSNPCHEDAICDTNPVSGRAICTCPPGFTGGACDQ	392
Qy	224	-----LCPPGKHGPQCE-----	235
Db	393	DVDECSIGANPCEHLGRCVNTQGSFLCQCGRGYTGPRCETDVNECLSGPCRNPQATCLDRI	452
Qy	236	-----QRCPCQNGGVC--HHVTG--ECSCPSGWMGTVC-----	264
Db	453	GQFTCICMAGFTGTGFCEVDIDECQSSPCVNGGVCKDRVNGFSCTCPSGFSGSTCQLDVDE	512
Qy	265	-----GQP-----CPEGRFGKNCSQ---ECQ---CHNGGTCD--ATGQCHC	298
Db	513	CASTPCRNGAKCVDQPDGYECRAEGFEGTLCERNVDDCSPDPCHHGRCVDGIASFSCAC	572
Qy	299	SPGYTGERCQDE-----CPVGTYGVLG-----AETCQ	325
Db	573	APGYTGIRCESQVDECRSQPCRYGGKCLDLVDKYLRCPPGTTGVNCEVNIDDCASNPT	632
Qy	326	---CVNGGKCYHVSAGLCEAGFAGERCEARL-----CPEGLYGIKCDKRCP	369
Db	633	FGVCRDGINRYD----CVCQPGFTGPLCNVEINECASSPCGEGGSCVDGENGFHC--LCP	686
Qy	370	-----CHLENTHS--CHPMG--ECACKPGWSGLYCNE-----	398
Db	687	PGSLPPLCLPANHPCAHKPCSHGVCHDAPGGFQVCVDPGWSGPRCSQSLAPDACESQPCQ	746
Qy	399	-----TCSPGFYGEACQOI--CS---CQNGADCDSVTGK---CTCAPGFKG	436
Db	747	AGGTCTSDGIGFHCTCAPGFQGHQCEVLSPTPSLCEHGGHCESDPDQLTVCSCPPGWQG	806
Qy	437	IDC-----STPC--PLGT--YGINCSSRCGCKNDV-----CSP-----	466
Db	807	PRCQQDVDEECAGASPCGPHGTCTNLPGSFRCICHGGYTGPFCDDIDDCDPNPCLNGGSC	866
Qy	467	VDG----SCTCKAGWHGVDC-----SIRCPSGTWGFGCN--	496
Db	867	QDGVGSFSCSCLSGFAGPRCARDVDECLSSPCGPGTCTDHVASFTCTCPPGYGGFHCETD	926
Qy	497	-LTC---QCLNGGACNTLDG---TCTCAPGWRGEKC-----	525
Db	927	LLDCSPSSCFNGGTC--VDGVNSFSCLCRPGYTGTHCQYKVDPCFSRPLHGGICNPHTS	984
Qy	526	--ELPCQDGTYGILNCAERCD-----CSHADGCHPTTGHCRCLPGWSGVHCD-----	569
Db	985	GFECTCREGFTGNQCQNPVDWCSQAPCQNGGRCVQTGAYCICPPEWSGPLCDIPSLPCTE	1044
Qy	570	-----SVCAEGRWGPNCSL---PC	585

RESULT 11

TENX_HUMAN

ID TENX_HUMAN STANDARD; PRT; 4289 AA.
AC P22105; P78530; P78531; Q08424; Q9UMG7;
DT 01-AUG-1991 (Rel. 19, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Tenascin X precursor (TN-X) (Hexabrachion-like).
GN TNXB OR TNX OR XB OR HXBL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Rowen L., Dankers C., Baskin D., Faust J., Loretz C., Ahearn M.E.,
RA Banta A., Schwartzell S., Smith T.M., Spies T., Hood L.;
RT "Sequence determination of 300 kilobases of the human class III MHC
RT locus.";
RL Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE OF 1-747 AND 1687-1944 FROM N.A.
RC TISSUE=Leukocyte;
RX MEDLINE=93300909; PubMed=7686164;
RA Bristow J., Tee M.K., Gitelman S.E., Mellon S.H., Miller W.L.;
RT "Tenascin-X: a novel extracellular matrix protein encoded by the human
RT XB gene overlapping P450c21B.";
RL J. Cell Biol. 122:265-278(1993).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM XB-SHORT).
RC TISSUE=Adrenal gland;
RX MEDLINE=96015044; PubMed=8530023;
RA Tee M.K., Thomson A.A., Bristow J., Miller W.L.;
RT "Sequences promoting the transcription of the human XA gene
RT overlapping P450c21A correctly predict the presence of a novel,
RT adrenal-specific, truncated form of tenascin-X.";
RL Genomics 28:171-178(1995).
RN [4]
RP SEQUENCE OF 1-23 FROM N.A.
RC TISSUE=Fetal adrenal gland;
RX MEDLINE=97081760; PubMed=8923003;
RA Speek M., Barry F., Miller W.L.;
RT "Alternate promoters and alternate splicing of human tenascin-X, a
RT gene with 5' and 3' ends buried in other genes.";
RL Hum. Mol. Genet. 5:1749-1758(1996).
RN [5]
RP SEQUENCE OF 3470-4289 FROM N.A.
RX MEDLINE=89367293; PubMed=2475872;
RA Morel Y., Bristow J., Gitelman S.E., Miller W.L.;
RT "Transcript encoded on the opposite strand of the human steroid 21-
RT hydroxylase/complement component C4 gene locus.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:6582-6586(1989).
RN [6]
RP DISEASE.

RX MEDLINE=21468843; PubMed=11642233;
 RA Schalkwijk J., Zweers M.C., Steijlen P.M., Dean W.B., Taylor G.,
 RA van Vlijmen I.M., van Haren B., Miller W.L., Bristow J.;
 RT "A recessive form of the Ehlers-Danlos syndrome caused by tenascin-X
 RT deficiency.";
 RL New Engl. J. Med. 345:1167-1175(2001).
 CC -!- FUNCTION: Appears to mediate interactions between cells and the
 CC extracellular matrix. Substrate-adhesion molecule that appears to
 CC inhibit cell migration. May play a role in supporting the growth
 CC of epithelial tumors.
 CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Comment=Additional isoforms seem to exist;
 CC Name=XB;
 CC IsoId=P22105-1; Sequence=Displayed;
 CC Name=XB-short;
 CC IsoId=P22105-2; Sequence=VSP_001418;
 CC -!- TISSUE SPECIFICITY: Highly expressed in fetal adrenal, in fetal
 CC testis, fetal smooth, striated and cardiac muscle. Isoform XB-
 CC short is only expressed in the adrenal gland.
 CC -!- DISEASE: Association with congenital adrenal hyperplasia.
 CC -!- DISEASE: Defects in TNXB are the cause of Ehlers-Danlos-like
 CC syndrome [MIM:606408]. This clinically distinct form of Ehlers-
 CC Danlos syndrome is characterized by hyperextensible skin,
 CC hypermobile joints, and tissue fragility, but it lacks atrophic
 CC scars and delayed wound healing. Inheritance is autosomal
 CC recessive.
 CC -!- SIMILARITY: Contains 19 EGF-like domains.
 CC -!- SIMILARITY: Contains 32 fibronectin type III domains.
 CC -!- SIMILARITY: Contains 1 fibrinogen C-terminal domain.
 CC -!- CAUTION: There are two genes for TN-X: TNXA and TNXB. TNXA is a
 CC partial gene which can sometimes recombine with TNXB.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U89337; AAB47488.1; -.
 DR EMBL; AF019413; AAB67981.1; -.
 DR EMBL; X71923; CAA50739.1; -.
 DR EMBL; Y13782; CAA74109.1; -.
 DR EMBL; Y13783; CAA74110.1; -.
 DR EMBL; U24488; AAB41287.1; -.
 DR EMBL; U52696; AAC50889.1; -.
 DR EMBL; M25813; AAA35884.1; -.
 DR PIR; A40701; A40701.
 DR HSSP; P02671; 1FZD.
 DR Genew; HGNC:11976; TNXB.
 DR MIM; 600985; -.
 DR MIM; 606408; -.
 DR GO; GO:0005578; C:extracellular matrix; NAS.
 DR GO; GO:0007160; P:cell-matrix adhesion; NAS.

DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002181; Fibrinogen_C.
 DR InterPro; IPR008957; FN_III-like.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00008; EGF; 8.
 DR Pfam; PF00147; fibrinogen_C; 1.
 DR Pfam; PF00041; fn3; 33.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00181; EGF; 7.
 DR SMART; SM00186; FBG; 1.
 DR SMART; SM00060; FN3; 30.
 DR PROSITE; PS00022; EGF_1; 18.
 DR PROSITE; PS01186; EGF_2; 19.
 DR PROSITE; PS50026; EGF_3; 8.
 DR PROSITE; PS00514; FIBRIN_AG_C_DOMAIN; 1.
 KW Glycoprotein; Cell adhesion; Repeat; EGF-like domain; Coiled coil;
 KW Extracellular matrix; Alternative splicing; Signal;
 KW Ehlers-Danlos syndrome.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 4289 TENASCIN X.
 FT DOMAIN 156 168 EGF-LIKE 1 (INCOMPLETE).
 FT DOMAIN 183 213 EGF-LIKE 2.
 FT DOMAIN 214 244 EGF-LIKE 3.
 FT DOMAIN 245 275 EGF-LIKE 4.
 FT DOMAIN 276 306 EGF-LIKE 5.
 FT DOMAIN 307 337 EGF-LIKE 6.
 FT DOMAIN 338 368 EGF-LIKE 7.
 FT DOMAIN 369 399 EGF-LIKE 8.
 FT DOMAIN 400 430 EGF-LIKE 9.
 FT DOMAIN 431 461 EGF-LIKE 10.
 FT DOMAIN 462 492 EGF-LIKE 11.
 FT DOMAIN 493 523 EGF-LIKE 12.
 FT DOMAIN 524 554 EGF-LIKE 13.
 FT DOMAIN 555 585 EGF-LIKE 14.
 FT DOMAIN 586 616 EGF-LIKE 15.
 FT DOMAIN 617 647 EGF-LIKE 16.
 FT DOMAIN 648 679 EGF-LIKE 17.
 FT DOMAIN 684 714 EGF-LIKE 18.
 FT DOMAIN 715 746 EGF-LIKE 19.
 FT DOMAIN 792 872 FIBRONECTIN TYPE-III 1.
 FT DOMAIN 901 922 COILED COIL (POTENTIAL).
 FT DOMAIN 941 1021 FIBRONECTIN TYPE-III 2.
 FT DOMAIN 1047 1127 FIBRONECTIN TYPE-III 3.
 FT DOMAIN 1149 1226 FIBRONECTIN TYPE-III 4.
 FT DOMAIN 1246 1327 FIBRONECTIN TYPE-III 5.
 FT DOMAIN 1348 1429 FIBRONECTIN TYPE-III 6.
 FT DOMAIN 1459 1540 FIBRONECTIN TYPE-III 7.
 FT DOMAIN 1561 1642 FIBRONECTIN TYPE-III 8.
 FT DOMAIN 1659 1736 FIBRONECTIN TYPE-III 9.
 FT DOMAIN 1756 1836 FIBRONECTIN TYPE-III 10.
 FT DOMAIN 1856 1939 FIBRONECTIN TYPE-III 11.
 FT DOMAIN 1962 2039 FIBRONECTIN TYPE-III 12.
 FT DOMAIN 2069 2150 FIBRONECTIN TYPE-III 13.
 FT DOMAIN 2167 2248 FIBRONECTIN TYPE-III 14.
 FT DOMAIN 2266 2347 FIBRONECTIN TYPE-III 15.

FT	DOMAIN	2365	2446	FIBRONECTIN TYPE-III 16.
FT	DOMAIN	2471	2552	FIBRONECTIN TYPE-III 17.
FT	DOMAIN	2582	2663	FIBRONECTIN TYPE-III 18.
FT	DOMAIN	2688	2769	FIBRONECTIN TYPE-III 19.
FT	DOMAIN	2794	2875	FIBRONECTIN TYPE-III 20.
FT	DOMAIN	2889	2972	FIBRONECTIN TYPE-III 21.
FT	DOMAIN	2997	3078	FIBRONECTIN TYPE-III 22.
FT	DOMAIN	3105	3186	FIBRONECTIN TYPE-III 23.
FT	DOMAIN	3211	3292	FIBRONECTIN TYPE-III 24.
FT	DOMAIN	3307	3384	FIBRONECTIN TYPE-III 25.
FT	DOMAIN	3399	3481	FIBRONECTIN TYPE-III 26.
FT	DOMAIN	3494	3575	FIBRONECTIN TYPE-III 27.
FT	DOMAIN	3601	3682	FIBRONECTIN TYPE-III 28.
FT	DOMAIN	3699	3787	FIBRONECTIN TYPE-III 29.
FT	DOMAIN	3801	3879	FIBRONECTIN TYPE-III 30.
FT	DOMAIN	3890	3971	FIBRONECTIN TYPE-III 31.
FT	DOMAIN	3978	4059	FIBRONECTIN TYPE-III 32.
FT	DOMAIN	4071	4289	FIBRINOGEN C-TERMINAL.
FT	SITE	1748	1750	CELL ATTACHMENT SITE (POTENTIAL).
FT	DISULFID	187	197	BY SIMILARITY.
FT	DISULFID	191	202	BY SIMILARITY.
FT	DISULFID	204	213	BY SIMILARITY.
FT	DISULFID	218	228	BY SIMILARITY.
FT	DISULFID	222	233	BY SIMILARITY.
FT	DISULFID	235	244	BY SIMILARITY.
FT	DISULFID	249	259	BY SIMILARITY.
FT	DISULFID	253	264	BY SIMILARITY.
FT	DISULFID	266	275	BY SIMILARITY.
FT	DISULFID	280	290	BY SIMILARITY.
FT	DISULFID	284	295	BY SIMILARITY.
FT	DISULFID	297	306	BY SIMILARITY.
FT	DISULFID	311	321	BY SIMILARITY.
FT	DISULFID	315	326	BY SIMILARITY.
FT	DISULFID	328	337	BY SIMILARITY.

Query Match 18.9%; Score 682; DB 1; Length 4289;

Best Local Similarity 28.4%; Pred. No. 1.3e-36;

Matches 191; Conservative 41; Mismatches 178; Indels 262; Gaps 38;

Qy	94	CCPGFYESG-----EMCVPHCADKCVHG-----RCIAPNTCQCEPGWGGTNCSSACDG	141
		: : :	
Db	125	CCPASAQAGTGQTDVRTLC-----SLHGVFDLSRC-----TCSCEPGWGGPTCSDDPTDA	173
Qy	142	D-----HWGPHC---TSRCQCKNGALCNPTG	165
		: : :	
Db	174	EIPPSPPSASGSCPDNDQGRGVRGRCVCFPGYTGPSGWPSCPGDCQGRGRC--VQG	231
Qy	166	ACHCAAGFRGWRCEDR-CEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFD-ED	223
		: : :	
Db	232	VCVCRAGFSGPDCSQRSCPRG-----CSQRGRCEG-----GRCVCDPGYTGDDCGMR	278
Qy	224	LCPPGKHGPQCEQRCPCQNGGVCHHVTGECSCPSGWMGTVCG-QPCPEGRFGKNCSQECQ	282
		: : :	
Db	279	SCPRG-----CSQRGRCEG-----GRCVCNPGYTGEDCGVRSCPRG-----CSQRGR	320
Qy	283	CHNGGTCAATGQCHCSPGYTGERC--QDECPVGTYGVLCAETCQCVNGGKCYHVSAGCLC	341
		: : : :	

Db 321 CKD-----GRCVCDPGYTGEDCGTRSCP-----WDCGEGGRC--VDGRCVC 359
 QY 342 EAGFAGERCEARLCPE-----GLYGIKCDKR-CPCHLENTHSCHPMSG 383
 | : | | | | | | | | | | | | | | | |
 Db 360 WPGYTGEDCSTRTCPDRCRGRGRCEEDGEICDTGYSGDDCGVRSCP GDCNQRGRCE--DG 417
 QY 384 ECACKPGWSGLYCNE-----TCSPGFYGEAC-QQIC--SCQNGADCD 422
 | | | | : | | | | | | | | | | | | | | | |
 Db 418 RCVCWPGYTGTD CGSRACPRDCRGRGRCENGVCVCNAGYSGEDCGVRSCP GDCRGRGRCE 477
 QY 423 SVTGKCTCAPGFKGIDCST-----PCPLGTYGINCSS-RC--GCKND 461
 | | : | | | | | | | | | | | | | | | |
 Db 478 S--GRCMCWPGYTGRDCGTRACPGDCRGRGRGVDCRVCNPGFTGEDCGSRRCPGDCRGH 535
 QY 462 AVCSPVDGSCTCKAGWHGVDCSIR-CPSGTWGFGCNLTCQCLNG----- 504
 : | | | | | | | | | | | | | | | | | | | |
 Db 536 GLCE--DGVCCVDAGYSGEDCSTRSCP GGCRCRG-----QCLDGRCVCEDGYSGEDCGVR 588
 QY 505 -----GACNTLDGTCTCAPGWRGEKCELP-----CQDGTYGLN 537
 | | | | | | | | | | | | | | | | | | | |
 Db 589 QCPNDCSQHGVQC--DGVCI CWEGYVSEDCSIRTCPSNCHGRGRCEEGRCLCDPGYTGP 646
 QY 538 CAER---CDCSHADGCHPTTGHCRCCLPGWSGVHC-----DSV 571
 | | | | | | | | | | | | | | | | | | | |
 Db 647 CATRMCPADCRGRGRG--VQGVCLCHVGYGGEDCGQEEPPASACPGGCGPRELCRAGQCV 704
 QY 572 CAEGRWGPNC SL 583
 | | | | : : : :
 Db 705 CVEGFRGPDCAI 716

RESULT 12

NTC2_MOUSE

ID NTC2_MOUSE STANDARD; PRT; 2470 AA.
 AC O35516; Q06008; Q60941;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 2 precursor (Notch 2) (Notch
 DE B).
 GN NOTCH2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Thymus;
 RA Hamada Y., Higuchi M., Tsujimoto Y.;
 RT "Complete amino acid sequence and mutliform transcripts encoded by a
 RT single copy of mouse Notch2 gene."
 RL Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE OF 316-1518 FROM N.A.
 RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
 RX MEDLINE=93178563; PubMed=8440332;
 RA Lardelli M., Lendahl U.;

RT "Motch A and Motch B-two mouse Notch homologues coexpressed in a
 RT wide variety of tissues.";
 RL Exp. Cell Res. 204:364-372(1993).
 RN [3]
 RP SEQUENCE OF 1765-2153 FROM N.A.
 RX MEDLINE=97075110; PubMed=8917536;
 RA Milner L.A., Bigas A., Kopan R., Brashem-Stein C., Bernstein I.D.,
 RA Martin D.I.;
 RT "Inhibition of granulocytic differentiation by mNotch1.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:13014-13019(1996).
 RN [4]
 RP FUNCTION.
 RX MEDLINE=99396706; PubMed=10393120;
 RA Hamada Y., Kadokawa Y., Okabe M., Ikawa M., Coleman J.R.,
 RA Tsujimoto Y.;
 RT "Mutation in ankyrin repeats of the mouse Notch2 gene induces early
 RT embryonic lethality.";
 RL Development 126:3415-3424(1999).
 RN [5]
 RP DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.
 RX MEDLINE=95333893; PubMed=7609614;
 RA Higuchi M., Kiyama H., Hayakawa T., Hamada Y., Tsujimoto Y.;
 RT "Differential expression of Notch1 and Notch2 in developing and adult
 RT mouse brain.";
 RL Brain Res. Mol. Brain Res. 29:263-272(1995).
 RN [6]
 RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
 RX MEDLINE=21523956; PubMed=11518718;
 RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
 RT "Murine notch homologs (N1-4) undergo presenilin-dependent
 RT proteolysis.";
 RL J. Biol. Chem. 276:40268-40273(2001).
 RN [7]
 RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
 RX MEDLINE=21374376; PubMed=11459941;
 RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
 RT "Conservation of the biochemical mechanisms of signal transduction
 RT among mammalian Notch family members.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity). May play an essential role in
 CC postimplantation development, probably in some aspect of cell
 CC specification and/or differentiation.
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=O35516-1; Sequence=Displayed;

```

CC      Name=2;
CC      IsoId=O35516-2; Sequence=VSP_001405;
CC      Note=No experimental confirmation available;
CC      -!- TISSUE SPECIFICITY: Expressed in the brain, liver, kidney,
CC          neuroepithelia, somites, optic vesicles and branchial arches, but
CC          not heart.
CC      -!- DEVELOPMENTAL STAGE: Expressed in the embryonic ventricular zone,
CC          the postnatal ependymal cells, and the choroid plexus throughout
CC          embryonic and postnatal development.
CC      -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC          which is proteolytically cleaved by a furin-like convertase in the
CC          trans-Golgi network before it reaches the plasma membrane to yield
CC          an active, ligand-accessible form. Cleavage results in a C-
CC          terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC          ligand binding, it is cleaved by TNF-alpha converting enzyme
CC          (TACE) to yield a membrane-associated intermediate fragment called
CC          notch extracellular truncation (NEXT). This fragment is then
CC          cleaved by presenilin dependent gamma-secretase to release a
CC          notch-derived peptide containing the intracellular domain (NICD)
CC          from the membrane.
CC      -!- PTM: Phosphorylated.
CC      -!- SIMILARITY: Belongs to the NOTCH family.
CC      -!- SIMILARITY: Contains 35 EGF-like domains.
CC      -!- SIMILARITY: Contains 2 Lin/Notch repeats.
CC      -!- SIMILARITY: Contains 6 ANK repeats.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; D32210; BAA22094.1; -.
DR      EMBL; X68279; CAA48340.1; -.
DR      EMBL; U31881; AAC52924.1; -.
DR      PIR; A49175; A49175.
DR      HSSP; P16109; 1FSB.
DR      MGD; MGI:97364; Notch2.
DR      GO; GO:0005887; C:integral to plasma membrane; IC.
DR      GO; GO:0005515; F:protein binding; IPI.
DR      GO; GO:0002011; P:morphogenesis of an epithelial sheet; IMP.
DR      GO; GO:0007219; P:N signaling pathway; IC.
DR      InterPro; IPR002110; ANK.
DR      InterPro; IPR000152; Asx_hydroxyl_S.
DR      InterPro; IPR000742; EGF_2.
DR      InterPro; IPR001881; EGF_Ca.
DR      InterPro; IPR001438; EGF_II.
DR      InterPro; IPR006209; EGF_like.
DR      InterPro; IPR002049; Laminin_EGF.
DR      InterPro; IPR008297; Notch.
DR      InterPro; IPR000800; Notch_dom.
DR      Pfam; PF00023; ank; 6.
DR      Pfam; PF00008; EGF; 34.
DR      Pfam; PF00066; notch; 2.
DR      PIRSF; PIRSF002279; Notch; 1.

```

DR PRINTS; PR00010; EGF_BLOOD.
 DR PRINTS; PR00011; EGFLAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF_CA; 23.
 DR SMART; SM00004; NL; 3.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 22.
 DR PROSITE; PS00022; EGF_1; 33.
 DR PROSITE; PS01186; EGF_2; 27.
 DR PROSITE; PS50026; EGF_3; 35.
 DR PROSITE; PS01187; EGF_CA; 22.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation;
 KW Alternative splicing.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 2470 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.
 FT CHAIN 1666 2470 NOTCH EXTRACELLULAR TRUNCATION.
 FT CHAIN 1697 2470 NOTCH INTRACELLULAR DOMAIN.
 FT DOMAIN 26 1677 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1678 1698 POTENTIAL.
 FT DOMAIN 1699 2470 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 26 63 EGF-LIKE 1.
 FT DOMAIN 64 102 EGF-LIKE 2.
 FT DOMAIN 105 143 EGF-LIKE 3.
 FT DOMAIN 144 180 EGF-LIKE 4.
 FT DOMAIN 182 219 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 221 256 EGF-LIKE 6 (INCOMPLETE).
 FT DOMAIN 258 294 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 296 334 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 336 372 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 373 411 EGF-LIKE 10.
 FT DOMAIN 413 452 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 454 490 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 492 528 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 530 566 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 568 603 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 605 641 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 643 678 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 680 716 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 718 753 EGF-LIKE 19.
 FT DOMAIN 755 791 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 793 829 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 831 869 EGF-LIKE 22.
 FT DOMAIN 871 907 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 909 945 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 947 983 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 985 1021 EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1023 1059 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1061 1097 EGF-LIKE 28.
 FT DOMAIN 1099 1145 EGF-LIKE 29.
 FT DOMAIN 1147 1183 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1185 1221 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1223 1260 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1262 1300 EGF-LIKE 33.

FT	DOMAIN	1302	1345	EGF-LIKE 34.
FT	DOMAIN	1372	1410	EGF-LIKE 35.
FT	REPEAT	1418	1454	LIN/NOTCH 1.
FT	REPEAT	1501	1533	LIN/NOTCH 2.
FT	REPEAT	1825	1869	ANK 1.

Query Match 18.8%; Score 677; DB 1; Length 2470;
 Best Local Similarity 24.7%; Pred. No. 1.8e-36;
 Matches 220; Conservative 78; Mismatches 246; Indels 348; Gaps 56;

Qy	3	ISLNSCLSFICL----	LLCH-----	WIGTASPLNLE--	DPNVCSHW-----	ES	39	
Db	529	IDIDDCSSTPCLNGAKCIDHPNGYECQCATGFTGILCDENIDNCDPDPCHHGQCQDGIDS				588		
Qy	40	YSVTVQESYPHPF--	DQI--	YYTS-----	CTDILNWFKCT-----	RHRVSY----	76	
Db	589	YTCICNPGYMGAI	CSDQIDECYSSPCLNDGR	CIDLVNGYQCNCQPGTSGLNCEINFDDCA		648		
Qy	77	RTAYRHG--	EKTMYRRKSQCCPGF-----	YESGEMCV-----		106		
Db	649	SNPCM	HGVCVDGINRYSCVCS	PGFTGQRCNIDIDE	CASNPCRKGATCINDVNGFR	CICPE	708	
Qy	107	----	PHC-----	ADKCVHGRC---	IAPNTCQCEPGWGGTNCSSACDGDHWGPHCTSR		151	
Db	709	GPHHPSCYSQVNECLSNPCIHGNCTGGLSGYKCLCDAGWVGVNCE--	VDKN----	ECLSN		762		
Qy	152	CQCKNGALCNPITGA--	CHCAAGFRGWRCE---	DRC-----	EQGTYGNDCH-QRCQCQ-		198	
Db	763	-PCQNGGTCNNLVNGYRCTCKKGF	KGYNQVNIDE	CASNPCLNQGT	CFDDVSGYTCHCML		821	
Qy	199	--	NGATCDHVTGECRCPPGYTGAF	CED-----	LCPPGKHGPQCE---	QRC---	PCQ	241
Db	822	PYTGKNCQTVLAPCSPNPCENAAVCKEAPNFESF	SCLCAPGWQ	GKRCTVDVDE	CISKPCM		881	
Qy	242	NGGVCHHVTGE--	CSCPSGWMGTVC	GPCEGRFGKNC	SQEC---	QCHNGGTC--	DAATG	294
Db	882	NNGVCHNTQGSYVCECP	PGFSGMDCEEDI-----	NDCLANPCQNGGSCVDHVNTF		931		
Qy	295	QCHCSPGYTGERCQDE-----	CPVGT	YGVLC---	AETC--		324	
Db	932	SCQCHPGFIGDKCQ	TDMNECLSEPC	KNGGTCS	DYVNSYTCTCPAGFHGVHCEN	NIDECTE	991	
Qy	325	-QCVN	GGKCYHVS	G----	ACLCEAGFAG-----	ERCEAR-----	LC	355
Db	992	SSCFN	GGTC--	VDGINSF	SCLCPVGFTGPFCLHDINE	CSSNPCLNAGTCVDGLGTYRCIC	1049	
Qy	356	PEGLYG	IKCD-----	KRCPCHLE	NTHSCHPM	SGEACKPGWSGLYCNE-----	398	
Db	1050	PLGYTGKNCQTLVNLCS	RSRSPCKNKG	TCVQEKARPHCLCPPGWDGAYCDVLNV	SCKAAALQ		1109	
Qy	399	-----	TCS	PGFYGEACQO---	ICS---	CQNGADCDSVTG--	426	
Db	1110	KGVPVEHLCQHSGICINAGNTHHCQ	CLPGYTGSYCEEQLDE	CASNPCQH	GATCNDFIGGY		1169	
Qy	427	KCTCAPG	FKGIDCSTPCPLGTYGIN	CSSRCGCKND	AVCSPVDGSCTCKAGWHGVDCSIRC		486	
Db	1170	RCECVPGYQGVNCE-----	YEVDE	CQNQPCQNGGT	CIDL	VNHFKCS-----	C	1211

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Iasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oregon-R;
 RX MEDLINE=20196011; PubMed=10731137;
 RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
 RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,
 RA Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,
 RA Minana B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,
 RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,
 RA Modolell J., Peter A., Schoettler P., Werner M., Mourkioti F.,
 RA Beinert N., Dowe G., Schaefer U., Jaeckle H., Bucheton A.,
 RA Callister D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,
 RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,
 RA Glover D.M.;
 RT "From sequence to chromosome: the tip of the X chromosome of *D.*
 RT *melanogaster*.";
 RL Science 287:2220-2222(2000).
 RN [5]
 RP SEQUENCE OF 2505-2611 FROM N.A.
 RX MEDLINE=85099329; PubMed=2981631;
 RA Wharton K.A., Yedvobnick B., Finnerty V.G., Artavanis-Tsakonas S.;
 RT "opa: a novel family of transcribed repeats shared by the Notch locus
 RT and other developmentally regulated loci in *D. melanogaster*.";
 RL Cell 40:55-62(1985).
 RN [6]
 RP SEQUENCE OF 1-8 FROM N.A.
 RX MEDLINE=87257846; PubMed=3037327;

RA Kelley M.R., Kidd S., Berg R.L., Young M.W.;
 RT "Restriction of P-element insertions at the Notch locus of *Drosophila*
 RT *melanogaster*.";
 RL Mol. Cell. Biol. 7:1545-1548(1987).
 RN [7]
 RP INTERACTION WITH DX, AND MUTANT SU42C.
 RX MEDLINE=94215489; PubMed=8162848;
 RA Diederich R.J., Matsuno K., Hing H., Artavanis-Tsakonas S.;
 RT "Cytosolic interaction between *deltex* and Notch ankyrin repeats
 RT implicates *deltex* in the Notch signaling pathway.";
 RL Development 120:473-481(1994).
 RN [8]
 RP INTERACTION WITH DX.
 RX MEDLINE=95401878; PubMed=7671825;
 RA Matsuno K., Diederich R.J., Go M.J., Blaumueller C.M.,
 RA Artavanis-Tsakonas S.;
 RT "Deltex acts as a positive regulator of Notch signaling through
 RT interactions with the Notch ankyrin repeats.";
 RL Development 121:2633-2644(1995).
 RN [9]
 RP S3 CLEAVAGE BY PSN.
 RX MEDLINE=99221487; PubMed=10206646;
 RA Struhl G., Greenwald I.;
 RT "Presenilin is required for activity and nuclear access of Notch in
 RT *Drosophila*.";
 RL Nature 398:522-525(1999).
 RN [10]
 RP S3 CLEAVAGE BY PSN.
 RX MEDLINE=99221488; PubMed=10206647;
 RA Ye Y., Lukinova N., Fortini M.E.;
 RT "Neurogenic phenotypes and altered Notch processing in *Drosophila*
 RT Presenilin mutants.";
 RL Nature 398:525-529(1999).
 RN [11]
 RP S2 CLEAVAGE BY KUZ.
 RX MEDLINE=21657146; PubMed=11799064;
 RA Lieber T., Kidd S., Young M.W.;
 RT "kuzbanian-mediated cleavage of *Drosophila* Notch.";
 RL Genes Dev. 16:209-221(2002).
 RN [12]
 RP MUTANT MCD5.
 RX MEDLINE=21575956; PubMed=11719214;
 RA Ramain P., Khechumian K., Seugnet L., Arbogast N., Ackermann C.,
 RA Heitzler P.;
 RT "Novel Notch alleles reveal a *Deltex*-dependent pathway repressing
 RT neural fate.";
 RL Curr. Biol. 11:1729-1738(2001).
 RN [13]
 RP REVIEW.
 RX MEDLINE=22256570; PubMed=12369105;
 RA Portin P.;
 RT "General outlines of the molecular genetics of the Notch signalling
 RT pathway in *Drosophila melanogaster*: a review.";
 RL Hereditas 136:89-96(2002).
 CC -!- FUNCTION: Signaling protein, which regulates, with both positive
 CC and negative signals, the differentiation of at least central and
 CC peripheral nervous system and eye, wing disk, oogenesis, segmental

CC appendages such as antennae and legs, and muscles, through lateral
 CC inhibition or induction. Functions as a receptor for membrane-
 CC bound ligands Delta and Serrate to regulate cell-fate
 CC determination. Upon ligand activation, and releasing from the cell
 CC membrane, the Notch intracellular domain (NICD) forms a
 CC transcriptional activator complex with Su(H) (Suppressor of
 CC hairless) and activates genes of the E(spl) complex. Essential for
 CC proper differentiation of ectoderm.
 CC -!- SUBUNIT: Interacts with Su(H) when activated. Interacts with Dx
 CC via its ANK repeats.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Upon activation and
 CC S3 cleavage, it is released from the cell membrane and enters into
 CC the nucleus in conjunction with Su(H).
 CC -!- PTM: Upon binding its ligands such as Delta or Serrate, it is
 CC cleaved (S2 cleavage) in its extracellular domain, close to the
 CC transmembrane domain. S2 cleavage is probably mediated by Kuz. It
 CC is then cleaved (S3 cleavage) downstream of its transmembrane
 CC domain, releasing it from the cell membrane. S3 cleavage requires
 CC Psn.
 CC -!- SIMILARITY: Belongs to the NOTCH family.
 CC -!- SIMILARITY: Contains 36 EGF-like domains.
 CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -!- SIMILARITY: Contains 6 ANK repeats.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M16152; AAB59220.1; -.
 DR EMBL; M16153; AAB59220.1; JOINED.
 DR EMBL; M16149; AAB59220.1; JOINED.
 DR EMBL; M16150; AAB59220.1; JOINED.
 DR EMBL; M16151; AAB59220.1; JOINED.
 DR EMBL; K03508; AAA28725.1; -.
 DR EMBL; M13689; AAA28725.1; JOINED.
 DR EMBL; K03507; AAA28725.1; JOINED.
 DR EMBL; AE003426; AAF45848.2; -.
 DR EMBL; AL035436; CAB37610.1; -.
 DR EMBL; AL035395; CAB37610.1; JOINED.
 DR EMBL; M12175; AAA74496.1; -.
 DR EMBL; M16025; AAA28726.1; -.
 DR

Query Match 18.8%; Score 677; DB 1; Length 2703;
 Best Local Similarity 25.4%; Pred. No. 1.9e-36;
 Matches 208; Conservative 78; Mismatches 203; Indels 330; Gaps 51;

Qy 7 SCL----SFICLLLCHWIGTASPLNLED--PNVCSHWESYSVTVQESYPHPFDQIYYTSC 60
 ||| :| |: : : || :::: | | :|
 Db 502 SCLDDPGTFRVCVMPGFTGTQCEIDIDECQSNPC-----LNDGTC 541
 Qy 61 TDILNWFKCTRHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMC---VPHCADKCVHGR 117
 | :| ||| : | || :| | : | : |
 Db 542 HDKINGFKCS-----CALGF--TGARCQINIDDCQSQPCRN 576

Qy 118 -----CIAPNTCQCEPGWGGTNCS---SACDGDHWGPHCTSRCQCKNGALCNPITG-ACH 168
 || :||| ||: ||: | : || : | | : : |
 Db 577 GICHDSIAGYSCECPPGYTGTSCEININDCDSN-----PCHRGKCIDDVNSFKCL 626

Qy 169 CAAGFRGW-----RCEDR-----CEQGTYG-----NDCHQRCQ 196
 | | : | : | : || | : || | | : ||
 Db 627 CDPGYTGYICQKQINECESNPCQFDGHCQDRVGSYYCQCQAGTSGKNCEVNVNECHSN-P 685

Qy 197 CQNGATC-DHVTG-ECRCPPGYTGAFCEDLCPGKHPQCEQRCPCQNGGVC-HHVTG-E 252
 | |||| | : : ||| || || : : | || || | | :
 Db 686 CNNGATCIDGINSYKCQCVPGFTGQHCE-----KNVDECIS-SPCANNGVCIDQVNGYK 738

Qy 253 CSCPSGWMGTVC-----GQP-----CPEGRFGKNCS---QECQ-- 282
 | || | : | | || | || | ||
 Db 739 CECPRGFYDAHCLSDVDECASNPCVNEGRCEDGINEFICHCPPGYTGKRCELDIDECSSN 798

Qy 283 -CHNGGTC-DAATG-QCHCSPGYTGERCQ---DECPVGTYGVLCAETCQCVNGGKCY-HV 335
 | : |||| | | ||||| : || : | | || | | |
 Db 799 PCQHGGTCYDKLNAFSCQCMPPGYTGQKCETNIDDC-----VTNPCGNGGTCTIDKV 848

Qy 336 SG-ACLCEAGFAGERCEARLCPEGLYGIKC-DKRCPCHLENTHSCHPMSG---ECACKP 389
 : | | : | | || : : | | || : | | | | ||
 Db 849 NGYKCVCKVPFTGRDCESKMDP-----CASNRC---KNEAKCTPSSNFLDFSTCKL 897

Qy 390 GWSGLYCNE-----TCSPGFYGEAC---QQICS---CQN 417
 | : | || : | | : | | | | : |||
 Db 898 GYTGRYCEDEDIDECSLSSPCRNGASCLNVPGSYRCLCTKGYEGRDCAINTDDCASFPQN 957

Qy 418 GADCDSVTG--KCTCAPGFKGIDCST-----PCPLGTYGI 450
 | | | | | | | | |||| ||
 Db 958 GGTCLDGIGDYSLCVDGFDGKHCETDINECLSQPCQNGATCSQYVNSYTCTCPLGFSGI 1017

Qy 451 NCS-----SRCGCKNDAVCSVPDG---SCTCKAGWHGVDCSIR----- 485
 || : | | | : || : || | : | : | :
 Db 1018 NCQTNDEDCTESSCLNGGSC--IDGINGYNCSCLAGYSGANCQYKLNKCDSNPCLNGATC 1075

Qy 486 -----CPSGTWGFGCNL-----TCQCLNGGACNTL--DGTCTCAPGWRGEKCE- 526
 |||| | | : | || | : : : | : || | : | :
 Db 1076 HEQNNEYTCHCPSGFTGKQCSEYVDWCGQSPCENGATCSQMKHQFSCKCSAGWTGKLC DV 1135

Qy 527 --LPCQDGT--YGLNCAERCD---CSHADGCHPTTGHCRCCLPGWSGVHC----- 568
 : ||| || : : | | | | : || : |
 Db 1136 QTISCQDAADRKGLSLRQLCNNGTCKDYGNSHV----CYCSQGYAGSYCQKEIDECQSQP 1191

Qy 569 -----DSVCAEGRWGPNCNL-----PC 585
 : | : | | || | ||
 Db 1192 CQNGGTCDRLIGAYECQCRQGFGQNCNELNIDDCAPNPC 1230

RESULT 14

NTC4_HUMAN

ID NTC4_HUMAN STANDARD; PRT; 2003 AA.
 AC Q99466; O00306; Q99458; Q99940; Q9H3S8; Q9UII9; Q9UIJ0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Neurogenic locus notch homolog protein 4 precursor (Notch 4)
 DE (hNotch4).
 GN NOTCH4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND POLYMORPHISM OF POLY-LEU.
 RC TISSUE=Placenta;
 RX MEDLINE=97311416; PubMed=9168133;
 RA Sugaya K., Sasanuma S.-I., Nohata J., Kimura T., Fukagawa T.,
 RA Nakamura Y., Ando A., Inoko H., Ikemura T., Mita K.;
 RT "Gene organization of human NOTCH4 and (CTG)_n polymorphism in this
 RT human counterpart gene of mouse proto-oncogene Int3.";
 RL Gene 189:235-244(1997).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
 RC TISSUE=Bone marrow, and Heart;
 RX MEDLINE=98360091; PubMed=9693032;
 RA Li L., Huang G.M., Banta A.B., Deng Y., Smith T., Dong P.,
 RA Friedman C., Chen L., Trask B.J., Spies T., Rowen L., Hood L.;
 RT "Cloning, characterization, and the complete 56.8-kilobase DNA
 RT sequence of the human NOTCH4 gene.";
 RL Genomics 51:45-58(1998).
 RN [3]
 RP SEQUENCE OF 1-503 FROM N.A., AND VARIANTS GLN-117 AND GLN-317.
 RA Miyagawa T., Tokunaga K., Hojho H.;
 RT "Human notch4 gene variant.";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
 RN [4]
 RP IDENTIFICATION OF LIGANDS.
 RX MEDLINE=99180765; PubMed=10079256;
 RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
 RA Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
 RT "Human ligands of the Notch receptor.";
 RL Am. J. Pathol. 154:785-794(1999).
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs. May regulate branching morphogenesis in the
 CC developing vascular system (By similarity).
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Comment=Experimental confirmation may be lacking for some
 CC isoforms;
 CC Name=1;
 CC IsoId=Q99466-1; Sequence=Displayed;
 CC Name=2;

CC IsoId=Q99466-2; Sequence=VSP_001406;
CC Name=3;
CC IsoId=Q99466-3; Sequence=VSP_001407;
CC -!- TISSUE SPECIFICITY: Highly expressed in the heart, moderately in
CC the lung and placenta and at low levels in the liver, skeletal
CC muscle, kidney, pancreas, spleen, lymph node, thymus, bone marrow
CC and fetal liver. No expression was seen in adult brain or
CC peripheral blood leukocytes.
CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-Golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane (By similarity).
CC -!- PTM: Phosphorylated (By similarity).
CC -!- POLYMORPHISM: The poly-Leu region of NOTCH4 (in the signal
CC peptide) is polymorphic and the number of Leu varies in the
CC population (from 6 to 12).
CC -!- SIMILARITY: Belongs to the NOTCH family.
CC -!- SIMILARITY: Contains 28 EGF-like domains.
CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.
CC -!- SIMILARITY: Contains 5 ANK repeats.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
CC in position 1438 to 1463.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; D63395; BAA09708.1; ALT_FRAME.
DR EMBL; D86566; BAA13116.1; -.
DR EMBL; U95299; AAC32288.1; -.
DR EMBL; U89335; AAC63097.1; -.
DR EMBL; AB023961; BAB20317.1; -.
DR EMBL; AB024520; BAA88951.1; -.
DR EMBL; AB024578; BAA88952.1; -.
DR HSSP; P08709; 1BF9.
DR Genew; HGNC:7884; NOTCH4.
DR MIM; 164951; -.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR008297; Notch.
DR InterPro; IPR000800; Notch_dom.

DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 26.
 DR Pfam; PF00066; notch; 2.
 DR PIRSF; PIRSF002279; Notch; 1.
 DR PRINTS; PR00010; EGFBLOOD.
 DR PRINTS; PR00011; EGFLAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 5.
 DR SMART; SM00179; EGF_CA; 11.
 DR SMART; SM00004; NL; 3.
 DR PROSITE; PS50297; ANK_REP_REGION; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 5.
 DR PROSITE; PS00010; ASX_HYDROXYL; 11.
 DR PROSITE; PS00022; EGF_1; 28.
 DR PROSITE; PS01186; EGF_2; 21.
 DR PROSITE; PS50026; EGF_3; 28.
 DR PROSITE; PS01187; EGF_CA; 9.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
 KW Triplet repeat expansion; Alternative splicing.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 2003 NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 4.
 FT CHAIN 1432 2003 NOTCH EXTRACELLULAR TRUNCATION
 FT (BY SIMILARITY).
 FT CHAIN 1467 2003 NOTCH INTRACELLULAR DOMAIN
 FT (BY SIMILARITY).
 FT DOMAIN 24 1447 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1448 1468 POTENTIAL.
 FT DOMAIN 1469 2003 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 24 63 EGF-LIKE 1.
 FT DOMAIN 64 115 EGF-LIKE 2.
 FT DOMAIN 118 155 EGF-LIKE 3.
 FT DOMAIN 156 192 EGF-LIKE 4.
 FT DOMAIN 194 232 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 234 274 EGF-LIKE 6.
 FT DOMAIN 276 312 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 314 353 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 355 391 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 392 430 EGF-LIKE 10.
 FT DOMAIN 432 473 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 475 511 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 513 549 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 551 587 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 589 625 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 626 659 EGF-LIKE 16.
 FT DOMAIN 661 689 EGF-LIKE 17.
 FT DOMAIN 691 727 EGF-LIKE 18.
 FT DOMAIN 729 765 EGF-LIKE 19.
 FT DOMAIN 767 803 EGF-LIKE 20.
 FT DOMAIN 806 842 EGF-LIKE 21.
 FT DOMAIN 844 880 EGF-LIKE 22.
 FT DOMAIN 882 928 EGF-LIKE 23.
 FT DOMAIN 930 966 EGF-LIKE 24.
 FT DOMAIN 968 1004 EGF-LIKE 25.
 FT DOMAIN 1006 1044 EGF-LIKE 26.
 FT DOMAIN 1046 1085 EGF-LIKE 27.

FT	DOMAIN	1087	1126	EGF-LIKE 28.
FT	DOMAIN	1130	1171	EGF-LIKE 29.
FT	DOMAIN	1472	1476	POLY-ARG.
FT	REPEAT	1165	1212	LIN/NOTCH 1.
FT	REPEAT	1213	1246	LIN/NOTCH 2.
FT	REPEAT	1247	1286	LIN/NOTCH 3.
FT	REPEAT	1633	1665	ANK 1.
FT	REPEAT	1666	1698	ANK 2.
FT	REPEAT	1700	1732	ANK 3.
FT	REPEAT	1733	1765	ANK 4.
FT	REPEAT	1766	1798	ANK 5.
FT	DISULFID	28	41	BY SIMILARITY.
FT	DISULFID	35	51	BY SIMILARITY.
FT	DISULFID	53	62	BY SIMILARITY.
FT	DISULFID	68	80	BY SIMILARITY.
FT	DISULFID	74	103	BY SIMILARITY.
FT	DISULFID	105	114	BY SIMILARITY.
FT	DISULFID	122	133	BY SIMILARITY.
FT	DISULFID	127	143	BY SIMILARITY.
FT	DISULFID	145	154	BY SIMILARITY.
FT	DISULFID	160	171	BY SIMILARITY.
FT	DISULFID	165	180	BY SIMILARITY.
FT	DISULFID	182	191	BY SIMILARITY.
FT	DISULFID	198	211	BY SIMILARITY.

Query Match 18.8%; Score 675.5; DB 1; Length 2003;
 Best Local Similarity 26.3%; Pred. No. 1.9e-36;
 Matches 211; Conservative 45; Mismatches 210; Indels 337; Gaps 46;

Qy	94	CCPGFYESGEMCVPHCADKC-----VHGRCIAPNT----CQCEPGWGGTNCSSACDGDH	143
		:	
Db	105	CLPGF--TGERCQAKLEDPCPPSFCSKRGRCHIQASGRPQCSCMPGWGTGEQCQLR-----	157
Qy	144	WGPHTSRCQCKNGALCNPITG--ACHCAAGFRGWRCE---DRCEQG----TYGNDCHQ-	193
		:: :	
Db	158	--DFCSAN-PCVNGGVCLATYPQIQCHCPPGFEGHACERDVNECFQDPGPCPKGTSCHNT	214
Qy	194	-----RCQ-----CQNGATC-----DHVTGECRCPPGYTGAFCE	222
		:	
Db	215	LGSFQCLCPVGQEGPRCEL RAGPCPPRGCSNGGTCQLMPEKDSTFHLCLCPPGFIGPDCE	274
Qy	223	-----DLCPPGKHG-----PQCEQRCP--CQNGGVCHH	248
Db	275	VNPDNCVSHQCQNGGTCQDGLDITYTCLCPETWTGWDCSEDVDECETQGPPHCRNGGTCQN	334
Qy	249	VTG--ECSCPSGWMGTVCQP-----CPEGFRFGKNCSQE-	280
Db	335	SAGSFHCVCVSGWGGTSCEENLDDCIAATCAPGSTCIDRVGSFSCLCPPGRTGLLCHLED	394
Qy	281	-C---QCHNGGTC--DAATGQ--CHCSPGYTGERCQ---DECPVGTYGVLCAETCQCVNG	329
Db	395	MCLSQPCHGDAQCSTNPLTGSTLCLCPGYSGPTCHQDLDECLMAQQG-----PSPCEHG	449
Qy	330	GKCYHVSGA--CLCEAGFAGERCEAR-----LCPEGLYGIK	363
		:	
Db	450	GSCLNTPGSFNCLCPPGYTGSRCEADHNECLSQPCHPGSTCLDLLATFHCLCPPGLEGQL	509

Qy	364	CD---KRC---PCHLENTHSCHPMSG---ECACKPGWSGLYCNE-----	398
		: :	
Db	510	CEVETNECASAPC--LNHADCHDLLNGFQCICLPGFSGTRCEEDIDECRSSPCANGGQCQ	567
Qy	399	-----TCSPGFYGEACQ---QIC---SCQNGADCDSVTGK--CTCAPGFKGIDCSTP	442
		:	
Db	568	DQPGAFHCKCLPGFEGPRCQTEVDECLSDPCPVGASCLDLPGAFFCLCPSGFTGQLCEVP	627
Qy	443	-----CPLGTYGI-----NCSSRCG-CKNDAVCSPVDGSCCTC	473
		: : :	
Db	628	LCAPNLCQPKQICKDQKDKANCLCPDGSPGCAPPEDNCTCHHGHCCQR-----SSVCV	679
Qy	474	KAGWHGVDC-----SIRCPSGTWGFGCN---LTCQ---CL	502
		: : :	
Db	680	DVGWTGPECEAELGGCISAPCAHGGTCYPQPSGYNCTCPTGYTGPTCSEEMTACHSGPCL	739
Qy	503	NGGACNTLDG--TCTCAPGWRGEKCE-----LPC-----QDGTYGLNCA-----	539
		: : :	
Db	740	NGGSCNPSPGGYYCTCPPSHTGPQCQTSTDYCVSAPCFNGGTCVNRPGTFSCLCAMGFQG	799
Qy	540	ERCD-----CSHADGCH--PTTGHCRCCLPGWSGVHCDS-----	570
		: : :	
Db	800	PRCEGKLRPSCADSPCRNRATCQDSPQGPRCLCPTGYTGGSCQTLMDLCAQKPCPRNSHC	859
Qy	571	-----VCAEGRWGPNCSLP	584
		: : :	
Db	860	LQTGPSFHCCLCQGWGTGPLCNLP	882

RESULT 15

NTC1 RAT

```

ID      NTC1_RAT          STANDARD;          PRT;   2531 AA.
AC      Q07008;
DT      01-NOV-1995 (Rel. 32, Created)
DT      15-JUL-1999 (Rel. 38, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Neurogenic locus notch homolog protein 1 precursor (Notch 1).
GN      NOTCH1.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Schwann cell;
RX      MEDLINE=92111383; PubMed=1764995;
RA      Weinmaster G., Roberts V.J., Lemke G.;
RT      "A homolog of Drosophila Notch expressed during mammalian
RT      development.";
RL      Development 113:199-205(1991).
RN      [2]
RP      REVISIONS TO 1652-1653.
RA      Weinmaster G.;
RL      Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN      [3]
RP      FUNCTION.
RX      MEDLINE=21094508; PubMed=11182080;

```

RA Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,
 RA Honjo T.;
 RT "Notch1 and Notch3 instructively restrict bFGF-responsive multipotent
 RT neural progenitor cells to an astroglial fate.";
 RL Neuron 29:45-55(2001).
 RN [4]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=93202015; PubMed=1295745;
 RA Weinmaster G., Roberts V.J., Lemke G.;
 RT "Notch2: a second mammalian Notch gene.";
 RL Development 116:931-941(1992).
 RN [5]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=21331789; PubMed=11438922;
 RA Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
 RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
 RT functional roles for the Notch-DSL signaling system during brain
 RT development.";
 RL J. Comp. Neurol. 436:167-181(2001).
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity). Acts instructively to control
 CC the cell fate determination of CNS multipotent progenitor cells,
 CC resulting in astroglial induction and neuron/oligodendrocyte
 CC suppression.
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus (By
 CC similarity).
 CC -!- TISSUE SPECIFICITY: Expressed in the brain, kidney and spleen.
 CC Expressed in postnatal central nervous system (CNS) germinal zones
 CC and, in early postnatal life, within numerous cells throughout the
 CC CNS. Found in both subventricular and ventricular germinal zones.
 CC -!- DEVELOPMENTAL STAGE: In the embryo, highest levels occur between
 CC days 12 and 14 and decrease rapidly to much lower levels in the
 CC adult.
 CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -!- PTM: Phosphorylated (By similarity).
 CC -!- SIMILARITY: Belongs to the NOTCH family.
 CC -!- SIMILARITY: Contains 36 EGF-like domains.
 CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.


```

CC      -!- SIMILARITY: Contains 5 ANK repeats.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X57405; CAA40667.1; -.
DR      HSSP; P00740; 1EDM.
DR      InterPro; IPR002110; ANK.
DR      InterPro; IPR000152; Asx_hydroxyl_S.
DR      InterPro; IPR000742; EGF_2.
DR      InterPro; IPR001881; EGF_Ca.
DR      InterPro; IPR001438; EGF_II.
DR      InterPro; IPR006209; EGF_like.
DR      InterPro; IPR002049; Laminin_EGF.
DR      InterPro; IPR008297; Notch.
DR      InterPro; IPR000800; Notch_dom.
DR      Pfam; PF00023; ank; 6.
DR      Pfam; PF00008; EGF; 35.
DR      Pfam; PF00066; notch; 3.
DR      PIRSF; PIRSF002279; Notch; 1.
DR      PRINTS; PR00010; EGFBLOOD.
DR      PRINTS; PR00011; EGFLAMININ.
DR      PRINTS; PR01452; NOTCH.
DR      SMART; SM00248; ANK; 6.
DR      SMART; SM00179; EGF_CA; 25.
DR      SMART; SM00004; NL; 2.
DR      PROSITE; PS50297; ANK_REP_REGION; 1.
DR      PROSITE; PS50088; ANK_REPEAT; 4.
DR      PROSITE; PS00010; ASX_HYDROXYL; 22.
DR      PROSITE; PS00022; EGF_1; 35.
DR      PROSITE; PS01186; EGF_2; 26.
DR      PROSITE; PS50026; EGF_3; 36.
DR      PROSITE; PS01187; EGF_CA; 21.
KW      Receptor; Transcription regulation; Activator; Differentiation;
KW      Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW      Transmembrane; Glycoprotein; Signal; Phosphorylation.
FT      SIGNAL          1      18      POTENTIAL.
FT      CHAIN           19     2531    NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT      CHAIN          1711    2531    NOTCH EXTRACELLULAR TRUNCATION (BY
FT                                     SIMILARITY).
FT      CHAIN          1744    2531    NOTCH INTRACELLULAR DOMAIN (BY
FT                                     SIMILARITY).
FT      DOMAIN          19     1723    EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM        1724    1746    POTENTIAL.
FT      DOMAIN          1747    2531    CYTOPLASMIC (POTENTIAL).
FT      DOMAIN          20      58     EGF-LIKE 1.
FT      DOMAIN          59      99     EGF-LIKE 2.
FT      DOMAIN         102     139     EGF-LIKE 3.
FT      DOMAIN         140     176     EGF-LIKE 4.
FT      DOMAIN         178     216     EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT      DOMAIN         218     255     EGF-LIKE 6.
FT      DOMAIN         257     293     EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).

```

FT	DOMAIN	295	333	EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	335	371	EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	372	410	EGF-LIKE 10.
FT	DOMAIN	412	450	EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	452	488	EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	490	526	EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	528	564	EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	566	601	EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	603	639	EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	641	676	EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	678	714	EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	716	751	EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	753	789	EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	791	827	EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	829	867	EGF-LIKE 22.
FT	DOMAIN	869	905	EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	907	943	EGF-LIKE 24.
FT	DOMAIN	945	981	EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	983	1019	EGF-LIKE 26.
FT	DOMAIN	1021	1057	EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1059	1095	EGF-LIKE 28.
FT	DOMAIN	1097	1143	EGF-LIKE 29.
FT	DOMAIN	1145	1181	EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1183	1219	EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1221	1265	EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1267	1305	EGF-LIKE 33.
FT	DOMAIN	1307	1346	EGF-LIKE 34.
FT	DOMAIN	1348	1384	EGF-LIKE 35.
FT	DOMAIN	1387	1426	EGF-LIKE 36.
FT	REPEAT	1445	1480	LIN/NOTCH 1.
FT	REPEAT	1481	1522	LIN/NOTCH 2.
FT	REPEAT	1523	1562	LIN/NOTCH 3.
FT	REPEAT	1917	1946	ANK 1.
FT	REPEAT	1950	1980	ANK 2.
FT	REPEAT	1984	2013	ANK 3.
FT	REPEAT	2017	2046	ANK 4.
FT	REPEAT	2050	2079	ANK 5.
FT	DOMAIN	1730	1733	POLY-ALA.
FT	DOMAIN	1891	1894	POLY-GLU.
FT	DOMAIN	2258	2261	POLY-PRO.
FT	DOMAIN	2497	2500	POLY-SER.
FT	SITE	1654	1655	CLEAVAGE BY (FURIN-LIKE PROTEASE) (BY
FT				SIMILARITY).
FT	DISULFID	24	37	BY SIMILARITY.
FT	DISULFID	31	46	BY SIMILARITY.
FT	DISULFID	48	57	BY SIMILARITY.
FT	DISULFID	63	74	BY SIMILARITY.
FT	DISULFID	68	87	BY SIMILARITY.
FT	DISULFID	89	98	BY SIMILARITY.
FT	DISULFID	106	117	BY SIMILARITY.
FT	DISULFID	111	127	BY SIMILARITY.
FT	DISULFID	129	138	BY SIMILARITY.
FT	DISULFID	144	155	BY SIMILARITY.
FT	DISULFID	149	164	BY SIMILARITY.
FT	DISULFID	166	175	BY SIMILARITY.
FT	DISULFID	182	195	BY SIMILARITY.
FT	DISULFID	189	204	BY SIMILARITY.

FT DISULFID 206 215 BY SIMILARITY.
 FT DISULFID 222 233 BY SIMILARITY.
 FT DISULFID 227 243 BY SIMILARITY.

Query Match 18.7%; Score 675; DB 1; Length 2531;
 Best Local Similarity 25.7%; Pred. No. 2.4e-36;
 Matches 208; Conservative 70; Mismatches 206; Indels 324; Gaps 48;

Qy 4 SLNSCLS-----FICLLLCHWIGTASPINLEDPNVCSHWESYSVTVQESY 48
 :| | | :| | | :| | | :|
 Db 603 NINECHSQPCRHHGGTCQDRDNYLCLCLKGTTGPNCEINLDD---CA----- 646

Qy 49 PHFPDQIYYTSCTDILNWFKCTRHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMC--- 105
 :| | :| | :| :| | | | :| :| | |
 Db 647 SNPCDS---GTCLDKIDGYECA-----CEPGY---TGSMCNVN 678

Qy 106 VPHCA-----DKCVHGRC---IAPNT 123
 : | | : | | | :
 Db 679 IDECAGSPCHNGGTCEdGIAGFTCRCPGYHDPTCLSEVNECNSNPCIHGACRDGLNGYK 738

Qy 124 CQCEPGWGGTNC---SSACDGDHWGPHCTSRCQCKNGALCNPTIG--ACHCAAGFRGWRC 178
 | | | | | | :| :| :| | | | :| | | | |
 Db 739 CDCAPGWSGTNCDINNNECESN-----PCVNGGTCKDMTSGYVCTCREGFSGPNC 788

Qy 179 EDRCEQGTYGNDCHQRCQCQNGATC-DHVTG-ECRCPPGYTGAFCEdLCPGKHGPQCEQ 236
 : | | :| | | | | :| | | | | :| |
 Db 789 Q-----TNINECASN-PCLNQGTClDDVAGYKCNCLPYTGATCEVVLAP-----C-A 834

Qy 237 RCPCQNGGVC---HHVTGECSCPSGWMGTVC-----GQPCPEGR----- 272
 ||:| ||| : : | ||:| | | || |
 Db 835 TSPCKNSGVCKESEDYESFSCVCPTGWQQTCEIDINECVKSPCRHGASCQNTNGSYRCL 894

Qy 273 -----FGKNCS---QECQ---CHNGGTCDAAATGQ--CHCSPGYTGERCQDE----- 310
 |:| | :| :| | | | | | | :| :| :| :|
 Db 895 CQAGYTGRNCESDIDDCRPNPCHNGGSCTDGVNAAFCDCLPGFQGAFCCEEDINECATNPC 954

Qy 311 -----CPVGTYGVLCAET-----CQCVNGGKCYHVSG-----ACLCEAG 344
 || | | :| | | | | | | | | |
 Db 955 QNGANCTDCVDSYTCTCPTGFNGIHCENNTPDCTESSCFNGGTC--VDGINSFTCLCPPG 1012

Qy 345 FAGERCEARLCPEGLYGI-KCDKRCPCHLENTHSCHPMMSG--ECACKPGWSGLYCNE--- 398
 | | | :| :| :| | | | | :| | | :| | |
 Db 1013 FTGSYCQ-----YDVNECDsr-PCLHGGT--CQDSYGTyKCTCPQGYTGLNCQNLVR 1061

Qy 399 --TCSPGFYGEACQQI-----CSCQN---GADCDsvTGKCTCAPGFKGIDCSTPCPLGTY 448
 :| | | | | :| :| :| :| :| :| :| :|
 Db 1062 WCDSAPCKNGGKCWQTNTQYHCECRSGWTGFNCdVLSVSCEVAAQKRGI DVTL LlcQHGGl 1121

Qy 449 GIN-----CSSRCG-----CKNDAVCSPVDG--SCTCKAGWHGVDCS 483
 : : | : | :| | | :| | | | :| :| :| :|
 Db 1122 CVDEEDKHCHCQAGYTGSYCEDEVDECSPNPCQNGATCTDYLGGFSCKCVAGYHGsnCS 1181

Qy 484 -----IRCPSGTWGFGCNLT---C-----QCLNG 504
 || || | | :| | :| |
 Db 1182 EEINECLSQCQNGGTClDLTNTYKCSCPRGtQGVHCEINVDdCHPPLDPASRS PKCFNN 1241

Qy 505 GACNTLDG--TCTCAPGWRGEKCE-----LPCQD-GTYGLNCAERCDCSHADGCHPT 553

					: :			: :	
Db	1242	GTCVDQVGGYTCTCPPGFVGERCEGDVNECLSNPCDPRGTQ--NCVQRVN-----							1289
Qy	554	TGHCRCLPGWSGVHCDSVCAEGRWGPNC 581							
				:	:				
Db	1290	DFHCECRAGHTGRRCESV-INGCRGKPC							1316

Search completed: March 26, 2004, 16:09:53
 Job time : 14.8459 secs